Towards engineering the microalga *Chlorella sorokiniana* for the production of tailored high-value oils

Xenia Spencer-Milnes

UCL (University College London)

A thesis submitted for the degree of Doctor of Philosophy (PhD)

September 2018

DECLARATION

I, Xenia Spencer-Milnes confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

……..

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor, Professor Saul Purton, for his continued support throughout these last four years and the opportunity to be part of such an interesting project. Also, many thanks go to my thesis committee and secondary supervisors: Dr Olga Sayanova for providing such expertise in the area of lipid metabolism and the opportunity to conduct some research at Rothamsted Research, Professor Kaila Srai for such constructive feedback, and Dr Vitor Pinheiro for providing invaluable support and advice through some difficult times.

I would also like to thank all members of the Algal Oils by Design sLoLa group for continued stimulating discussion and inspiration at meetings. I would especially like to thank Dr Mary Hamilton and Dr Richard Smith from Rothamstead Research for their patience and support in teaching me new techniques and putting up with a myriad of questions. I also must thank former lab members Noreen Hiegle for showing me the ropes using *Agrobacterium* and Dr Sofie Vonlanthen who established much *Chlorella* work in the lab and was very quick to respond to my flurry of email questions in the beginning.

Thanks is also due to the great bunch of people in the Purton lab who I have spent time with over the years and enjoyed getting up to mischief with at conferences and lab outings! Dr Rosie Young has been a great level head and provider of encouragement and advice during this process. Dr Laura Stoffels, Dr Priscilla Rajakumar, Dr Umaima Al Hoquani and Dr Alice Lui, Dr Max Blanshard and Dr Janet Waterhouse and Dr Fiona Li, thank you for all your help in the lab from knowing where things are to giving tips on protocols and solving little mis-haps along the way! To Saowalak Changko, Jing Cui and Lydia Mapstone, you've been great company and I wish you all the best with your continuing work. An enormous thanks is due to our lab manager Thushi Sivagnanam who keeps things running so smoothly and has so much patience for us.

Dr Henry Taunt has been a valuable source of knowledge over the years and an excellent provider of suitable beverages. I would especially like to say a massive

thank you to him for his incredible patience and encouragement during the final stages of the PhD. And to Juliana – we did it, woo! Thanks for being such a great partner in the lab and getting through the end together!

The most important thanks must go to my parents Sally and John, and brother Zed, whose love and support I know is unconditional. Without them I would be at a loss. Finally, I would like to thank Andy for being a rock through this process, and my housemates Uther, Sophie, Lauren and Alex for being such wonderful company, especially after a hard day.

ABSTRACT

This project explores the production of LC-PUFAs, such as the nutritionally relevant omega-3 fatty acids EPA and DHA, in the freshwater microalga *Chlorella sorokiniana* (UTEX 1230). *C. sorokiniana* is of interest due to its rapid growth rate, tolerance to high light and temperature, and ability to accumulate a large proportion of cell weight as lipids. Since wild-type *C. sorokiniana* terminates fatty acid synthesis at ALA (C18:3n3), a genetic engineering approach is required to produce LC-PUFAs.

Changes in FAME and lipid content in different growth conditions including carbon source, nitrogen source, and trophic state were assessed in this microalga. These investigations confirmed the common finding that nitrogen stress is a robust way to induce neutral lipid accumulation, but also highlighted the potential of pH change as an alternate stressor for TAG production.

To facilitate the heterologous gene expression needed to increase the range of fatty-acids produced by *C. sorokiniana*, this work develops and begins to characterise a toolbox of genetic parts for nuclear transformation. This combinatorial parts library, compatible with a standard cloning syntax to enable sharing between groups, comprises five coding sequences, four promoter/5'UTRs and four 3'UTR/terminators. The coding sequences include two antibiotic reporters, two fluorescent reporters and one lipid gene. Among the regulatory sequences, two were novel putative promoters from chlorovirus, as identified from detailed bioinformatics analysis of published transcriptomic and genomic data.

An attempt to improve an existing *Agrobacterium*-mediated transformation strategy was utilised to validate some of the parts within the library. Preliminary evidence suggests the putative chlorovirus promoters may be active within both *C. sorokiniana* and *Chlamydomonas reinhardtii*. Challenges with appropriate selectable markers and transformation efficiency highlight the difficulty in genetic engineering of microalgal strains, especially where these tools are not well established, but also showcase the pressing need for such foundational research to be conducted.

IMPACT STATEMENT

Microalgae are promising biotechnological platforms: lipid-rich biomass can be used for biofuels, and many species accumulate high-value products such as pigments and omega-3 oils, including EPA (20:5n-3) and DHA (22:6n-3), which are essential for human nutrition. However, there is often a trade-off in the production of these lipids, in that oil accumulation in the microalgae is stimulated when under stressful growth conditions such as nitrogen deprivation. The resulting decrease in growth rate and productivity can be a barrier to economic viability of industrial production. *Chlorella sorokiniana* UTEX 1230 is a small single-celled freshwater green alga of industrial relevance with regard to its high growth rate at warm temperatures and ability to accumulate lipids under optimised growth conditions.

The wider aim of this project is to explore the potential for production of omega-3 fatty acids in *C. sorokiniana*. By working towards providing an alternative source of the nutritionally essential omega-3 oils EPA and DHA, the output of this project may contribute to the reduction of the unsustainable pressure on the fishing industries. Additionally, western diets are typically associated with consumption of too much omega-6 compared to omega-3, therefore additional sources of these oils may help increase the health of the population.

The main outputs of this work relate to factors affecting lipid accumulation in *C. sorokiniana*, the development of a transformation methodology for this alga, and the creation of a library of shareable DNA parts for use in microalgae genetic engineering, including new putative regulatory sequences from chlorovirus genes. The study of lipid accumulation is important because it is relevant to improving process dynamics across several areas of the nascent algal biotech industry, including feed, food, and biofuels. Research conducted on the transformation methodology of *C. sorokiniana* informs and complements future efforts into this area. Finally, the parts library will allow for more rapid construction of algal expression cassettes in the Purton lab, and the use of a common syntax will encourage the sharing of parts across academic and industrial labs to facilitate future technological development. These

findings will be useful for both microalgal biotechnology industries and further academic research into basic physiology of microalgae.

ABBREVIATIONS

ACCase	acetyl-CoA carboxylase
ACP	acyl carrier protein
ALA	Alpha-Linoleic-Acid (18:3n3)
ARA	Arachidonic acid (20:4n6)
ATP	adenosine 5’-triphosphate
bp	basepair
CCAP	Culture Collection of Algae and Protozoa
cDNA	complementary deoxyribonucleic acid
CDS	Coding Sequence
CoA	coenzyme A
DAG	diacylglycerol
DGDG	digalactosyl diacylglycerol
DGTS	Diacylglyceryltrimethylhomoserine
DHA	docosahexaenoic acid (22:6n3)
DIG	digoxigenin-dUTP
DMSO	diemthyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	2’deoxynucleoside 5’-triphosphate
EDTA	ethylenediaminetetraacetic acid (disodium salt)
EPA	eicosapentaenoic acid (20:5n3)
FA	fatty acid
FAME	fatty acid methyl ester
FAS	Fatty acid synthase
GC	gas chromatography
GC-FID	gas-chromatography-flame ionisation detection
GC-MS	gas chromatography-mass spectrometry
GOI	gene of interest
GPAT	glycerol-3-phosphate acyltransferase
HPLC	high performance liquid chromatography
LA	linoleic acid (18:2n6)
LC-PUFA	long-chain poly-unsaturated fatty-acid
MGDG	monogalactosyl diacylglycerol
mRNA	messenger ribonucleic acid
MS	mass spectrometry
NAD(P)H	Nicotinamide adenine dinucleotide phosphate
nm	nanometer
NR	nile red
OD	optical density
PA	phosphatidic acid
PAT	poly-A tail

PC	phosphatidyl choline
PCR	polymerase chain reaction
PDAT	phospholipid diacylglycerol acyltransferase
PE	phosphatidylethanolamine
PEG	polyethylene glycol
PG	Phosphatidylglycerol
PI	phosphatidylinositol
PSI, PSII	photosystem I, photosystem II
PUFA	polyunsaturated fatty acid
RACE	Rapid amplification of cDNA ends
RBCS	ribulose-1,5-biphosphate carboxylase/oxygenase
rDNA	ribosomal deoxyribonucleic acid
RNA	ribonucleic acid
RNase	ribonuclease
rRNA	ribosomal ribonucleic acid
RT-PCR	reverse transcription-polymerase chain reaction
SD	standard deviation
SQDG	sulfoquinovosyl diacylglycerol
TAG	triacylglycerol
TAP	tris acetate phosphate
TE	tris EDTA
tris	(hydroxymethyl) aminomethane
UTEX	University of Texas
UTR	Untranslated Region
UV	ultraviolet
v/v	volume by volume
w/v	weight by volume
wt	wild-type

CONTENTS

FIGURES AND TABLES

CHAPTER 1 INTRODUCTION

1.1 What are microalgae

Microalgae are a diverse group of prokaryotic (cyanobacteria) or eukaryotic single-celled photosynthetic organisms which exist ubiquitously in freshwater or marine environments. The diversity of microalgae spans across kingdoms when including prokaryotic cyanobacteria, and even within the eukaryotes spans across many phyla and groups including: the red and green algae within the Archaeplastida; the diatoms, brown algae and Eustigmatophytes within the Stramenophiles (Heterokonts); Dinoflagellates within the Alveolates; and other more isolated groups such as the Haptophytes and Cryptophytes (**Figure 1-1**). This results in a wide diversity of metabolism, metabolites and life strategies which can be exploited for bioproducts in natural species, or used as a source of information on biosynthetic pathways for trans-expression in more standard work-horse production platform for biotechnology such as *E. coli* or yeast (Borowitzka, 2013; Hamed, 2016).

Microalgae are evolutionarily ancient (Falkowski et al., 2004; Leliaert et al., 2012, 2011). Although the precise date of their origin has been debated (Dalton, 2002), it was around 3 billion years ago when the first prokaryotic cyanobacteria ancestors originated, coinciding with the 'oxygen catastrophe': a huge shift in the earth's atmosphere where oxygenic photosynthesis caused it to become oxygenated which then facilitated the evolution of aerobic organisms (Clerck et al., 2012). Oxygenic photosynthesis was, and still is, a very successful life strategy and it is accepted that eukaryotic microalgae (and subsequently higher plants) evolved from the symbiotic engulfment of cyanobacteria by non-photosynthetic heterotroph organisms between 1 and 1.5 billion years ago. Over time the cyanobacterial partner lost its independence and became the plastid or chloroplast that we see today (Clerck et al., 2012; Falkowski et al., 2004; Leliaert et al., 2011). Further to this primary endosymbiosis, various subsequent gains, replacements and losses of plastids have occurred within the eukaryotic tree of life. Secondary or tertiary endosymbiosis events are proposed to explain this spread of plastids and the presence of extra

membranes surrounding the chloroplasts in algal groups such as the haptophytes and heterokonts (Stramenopiles) (Falkowski et al., 2004; Leliaert et al., 2012, 2011; Tirichine and Bowler, 2011). These events are proposed to have occurred convergently in different groups of eukaryotic heterotrophs, and involving different sources of the plastids (i.e. either green or red lineages) hence microalgae are not a monophyletic group (Andersen, 2013).

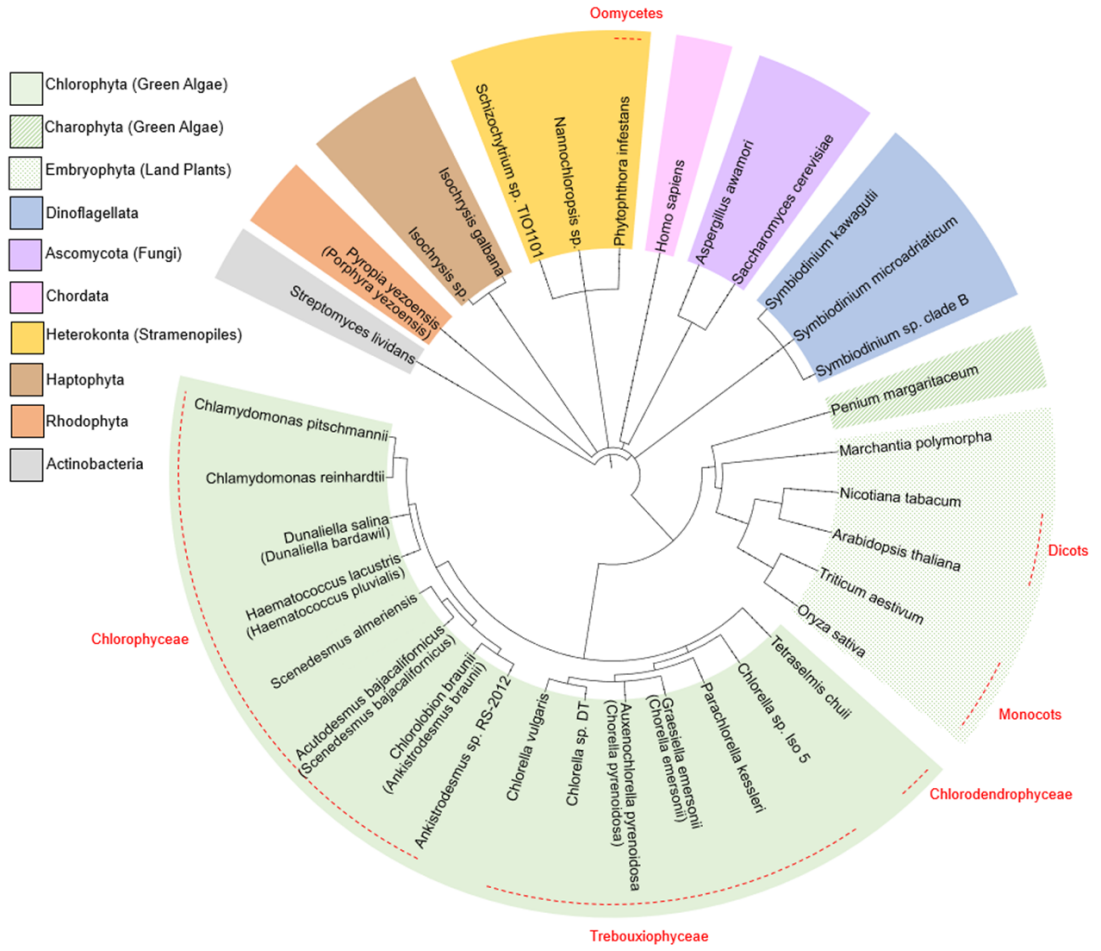


Figure 1-1 Microalgal diversity on the eukaryotic tree of life. Examples across common different groups are shown. Tree information was generated from NCBI taxonomy using the tool phyloT (available at http://phylot.biobyte.de/index.html) database version 2017.7 (biobyte solutions GmbH). Tree output was visualised using iTOL (Interactive tree of life, available at http://itol.embl.de/ (Letunic and Bork, 2016)).

Today the estimates of the biodiversity of microalgae are vast with over 40,000 species identified, but this is predicted to only be a fraction of the true number of species (Hu et al., 2008). This biodiversity has resulted in multiple different morphologies, trophic modes, motility mechanisms and life cycles (**Figure 1-2**). For example, *Volvox* is colonial, diatoms have silica shells of great variety and the most efficient form of Rubisco for photosynthesis in autotrophs, *Haematococcus pluvialis* and *Chlamydomonas reinhardtii* have both flagellate and non-flagellate forms depending on their environment or life cycle stage (Andersen, 2013; Fields and Kociolek, 2015). Microalgae have also spread out to occupy diverse niches, for example snow algae, acidophiles, marine, freshwater, and even growing on animal fur such as that of sloths (Andersen, 2013; Domozych et al., 2012). They are ecologically very important, for example the diatoms are thought to contribute up to 25 % of global primary productivity (Scala and Bowler, 2001), with eukaryotic phytoplankton as a whole estimated to contribute to over 45 % of Earth's net primary production ((Field et al., 1998) as cited in (Falkowski et al., 2004)).

Figure 1-2 Examples of the diversity of microalgal forms and habitats.

A: Symbiotic *Chlorella* within the cilliate *Paramecium bursaria* SAG 27.96 (https://sagdb.uni-goettingen.de). **B:** *Euglena gracilis* (Dr Ralf Wagner; http://www.dr-ralf-wagner.de).

C: Microalgal biomass on the head and back of a two-toed sloth (Pauli et al., 2014).

D: *Volvox carteri* adult multi-cellular spheroid (Matt and Umen, 2018). **E:** *Arthrospira fusiformis* (Spirulina) (Sili et al., 2012). **F:** Red snow algae (https://www.gettyimages.co.uk).

G: *Chlamydomonas reinhardtii* scanning electron micrograph (http://remf.dartmouth.edu).

H: Diatom shape diversity across multiple species. Microscopic (no scale provided) (https://commons.wikimedia.org/w/index.php?curid=5682386). **I:** Cyanobacterial toxic bloom in a Californian lake. (K. Macintyre, California Dpt of Fish and Wildlife, from: https://www.ecowatch.com/algae-blooms-climate-change-2017383600.html).

Examples of thousands of these species are maintained in culture collections around the world (Mata et al., 2010). These include the University of Coimbra (Portugal) with over 4000 strains and 1000 species, the SAG collection at Goettingen University (Germany) with over 2200 strains and over 1250 species, the University of Texas Algal Culture Collection (UTEX) in USA with ~2300 strains, the National Institute for Environmental Studies Collection (NIES) in Japan with 2150 strains and 700 species (Mata et al., 2010), and the Culture Collection of Algae and Protozoa at the Scottish Association for Marine Science in Scotland (CCAP) with 2570 strains and over 1290 species (SAMS, 2018). The maintenance of these strains is commonly carried out by sequential subculture in low-growth conditions, though increasingly cryopreservation is being explored to avoid genetic drift (Brand et al., 2013).

As seen above there is disparity between the numbers of 'strain' and 'species' and this is partly due to culture collections starting as early as the 1920s (Mata et al., 2010) when molecular phylogenetic techniques such as barcoding using the ITS-2 regions of rRNA (Coleman, 2003; Hall et al., 2010; Leliaert et al., 2014, 2012) were not available for taxonomy. As discussed later using the example of *Chlorella*, microalgal taxonomy is complicated by multiple historical methods of identification. However, culture collections routinely update their records with published reclassifications. Additionally, it is recognised that even those strains which show very similar genetic relationships via DNA 'barcodes' may be isolated from multiple places or at different times and have alternate physiologies and behaviours. Therefore, in culture collections this difference is recognised and designated as various strains of the same species. That even very closely related microalgae strains within the same species behave differently is a testament to their versatile metabolism. It is also a feature of microorganisms in general that evolutionary forces may act more quickly and in different ways than in higher organisms which can blur the boundaries of species (Brand et al., 2013; Leliaert et al., 2014, 2012).

The diversity of microalgae that is seen today is probably partly because they are so evolutionarily old combined with the fact that this single-celled photosynthetic life strategy was successful enough to have diversified across the whole kingdom of life during evolutionary history. Although this diversity is exciting and gives a large

number of untapped biosynthetic pathways and possible components (Gimpel et al., 2015; Guarnieri and Pienkos, 2015), in algal biotechnology it can lead to confusion as a result of drawing over-generalised conclusions from studies on microalgae from such diverse clades; for example the establishment of transformation protocols needs optimising at a species level (Gimpel et al., 2015).

1.2 Microalgae as a resource for bioproducts

Microalgae are an attractive biotechnological platform. They have high productivity compared to other terrestrial systems, giving 2-10 times more biomass per unit land area (Sayre, 2010). They also have the potential to act as bio-factories converting sunlight to useful products in a sustainable way without competing for land space like plant crops, and also fixing atmospheric carbon dioxide in the process (Guarnieri and Pienkos, 2015; Hamed, 2016; León-Bañares et al., 2004). Microalgal cells consist of the core components of protein, carbohydrate, lipids and nucleic acids – and also owing to their diversity, a wide array of other sometimes unique bioactive compounds such as antioxidants (such as carotenoids), phycobilins and sterols (Gimpel et al., 2015), therefore they are of interest for exploitation of these compounds as discussed below and summarised in **Figure 1-3**. Since this thesis is about engineering the lipids of *Chlorella* for a nutritional product, the discussion of microalgal uses below focuses on lipid-derived and nutritional products. There are, however, other important uses such as an expression platform for proteins, especially in the chloroplast which can act as a compartment to allow hyper-accumulation. For waste bioremediation they can be used for treatment of municipal organics (like sewage) or the absorption of specific toxins such as heavy metals like cadmium from mine leakage due to the ability of some species to absorb contaminants. These other uses are reviewed in (Delrue et al., 2016; Forján et al., 2015; Gangl et al., 2015; León-Bañares et al., 2004; Liu and Chen, 2016).

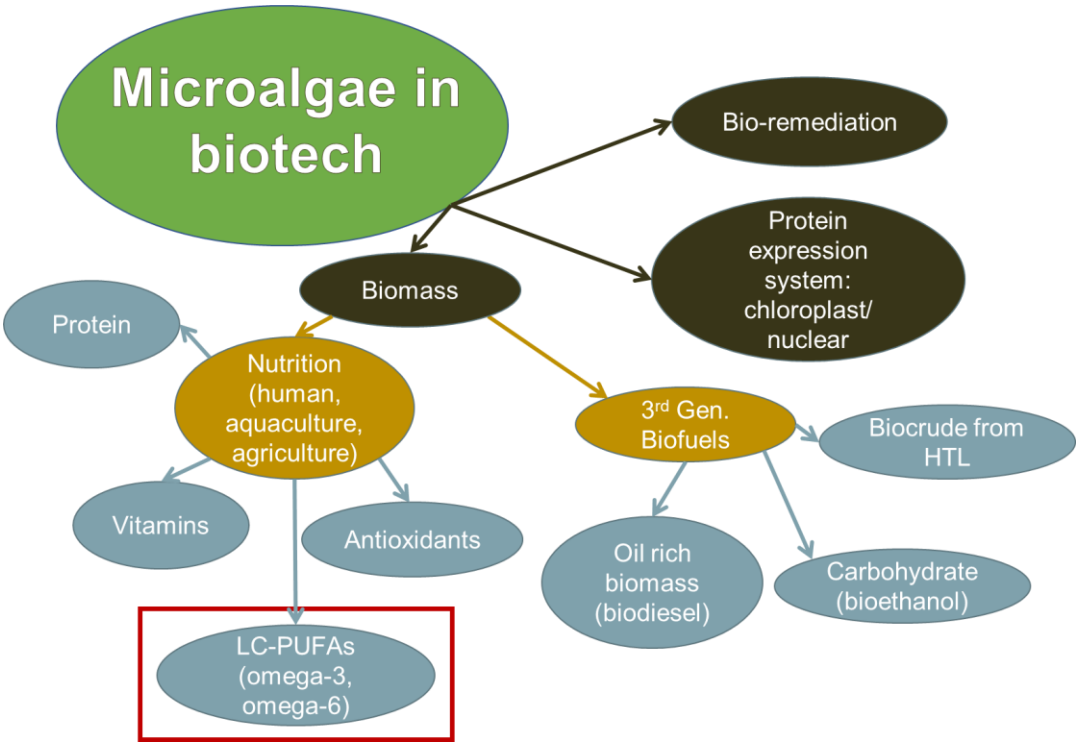


Figure 1-3 Key Biotechnological uses of microalgae with a focus on bioproducts. HTL: Hydrothermal Liquefaction. The background image is from http://eapsweb.mit.edu/news/2014/phytoplankton-diversity-versus-productivity-ocean.

1.2.1 Fatty acid nomenclature

Fatty acids are discussed extensively over the course of this work and can be described by multiple names. To avoid confusion, in this thesis nomenclature is in the form 18:3n3 which is (number of carbon atoms): (number of double bonds) n (position of the nearest double bond from the omega end of the fatty acid) (**Figure 1-4**).

ALA; α-Linolenic acid; (9Z,12Z,15Z)-9,12,15-Octadecatrienoic acid; 18:3Δ9,12,15; **18:3n3**

Figure 1-4 Nomenclature of fatty acids. The example is alpha-linoleic acid. The preferred nomenclature format used in this work is highlighted in bold, which counts the number of carbons from the omega (methyl) end of the chain to the double bond.

1.2.2 Biofuels from microalgae

1.2.2.1 Why does the world need biofuels?

The negative impacts of fossil fuel use in relation to the emission of greenhouse gasses such as carbon dioxide causing climate change are now well established, and with finite supplies of such fuels there is an acknowledged need for renewable alternatives (Hu et al., 2008; Mata et al., 2010; Sayre, 2010; Williams and Laurens, 2010). There are many alternative renewable alternatives such as solar, wind and tidal. However, there is still a large demand for renewable fuels with output as liquid fuel for cars (e.g. biodiesel), aviation, and other transport applications where hydrocarbon-cased liquid fuels are ultimately superior or essential due to the high energy density (Williams and Laurens, 2010). For example, the change of infrastructure to electric cars is a long way off, and there is a demand for renewables compatible with petroleum-based infrastructure (Sayre, 2010). Therefore, biofuels, that is liquid fuels (biodiesel to replace diesel, ethanol and other alcohols to replace gasoline, and bio-jet fuel to replace kerosine) derived from living biological sources rather than fossil ones, has had increasing interest (Mata et al., 2010).

1.2.2.2 The merits of microalgae as a source of biofuels

Microalgae were recognised as early as the 1940s as being rich in storage lipid droplets (composed of triacylglycerols or TAGs) depending on culturing conditions (Borowitzka, 2013; Hu et al., 2008). In the last few decades microalgae have been explored as a source of so-called 3rd generation biofuels; with 1st generation biofuels derived from plant oils or from sugars and starches fermented to ethanol, and 2nd generation bioethanol produced by breakdown and fermentation of plant lignocellulose (Naik et al., 2010). The advantages of microalgal biofuels over those derived from crop plants can include longer growing seasons, higher productivity, reduced water needs, less conflict of land space for food production and the ability to grow algae on waste stream inputs for combined bioremediation (Chen et al., 2015; Kiran et al., 2014; Maity et al., 2014; Rawat et al., 2013). A comparison of microalgal oil with other plant sources can be found in the review by Sajjadi *et al* (Sajjadi et al., 2018).

Ultimately the algal biomass is the source of the fuel or components for derivatising to fuel, and for some species it is also extraordinarily high in valuable components. The carbohydrate content of microalgae varies from 5-80% depending on species and consists of polysaccharides and starch which can be used for bioethanol production and are preferable compared to other plant sources due to reduced amount of contaminating lignocellulose (Tang et al., 2016a). For biodiesel and oil fuels, it is the lipid content of the microalgal cells that is relevant. Depending on the microalgal species and the cultivation strategy, microalgal lipid content can range from under 2 % up to 75 % (D'Alessandro and Antoniosi Filho, 2016). In general there are many algal species with superior accumulation qualities to the common plant stock such as oil-seed rape (Bellou et al., 2016; D'Alessandro and Antoniosi Filho, 2016; Minhas et al., 2016; Sajjadi et al., 2018; Schüler et al., 2017; Williams and Laurens, 2010; Zhan et al., 2017). Although this range of different content shows there are promising species, some authors argue that the diversity means it is hard to select the best species for commercial development (Fields and Kociolek, 2015).

TAGs (triacylglycerols) (**Figure 1-5**) are the preferred form of lipid for biofuel because the energy-rich acyl chains have a similar density to hydrocarbons, and they do not contain the phosphate or sulphate that would be considered impurities and would have a negative impact on the oil product (MacDougall et al., 2011; Williams and Laurens, 2010). The fatty acid chain length profile must also be considered because it contributes to the stability and combustion properties of the fuel. Although the precise mix is complex, mono-unsaturated fatty acids can be beneficial because they have a lower freezing point and therefore better cold-flow properties, but poly-unsaturated fatty acids can be undesirable because they have a negative impact on the cetane number (speed of combustion) and are more prone to oxidation, whereas saturated fatty acids provide the best cetane number (Olmstead et al., 2013; Ramos et al., 2009; Williams and Laurens, 2010). For chain length, a typical high quality fuel may consist of carbon chains of 18 in length, but shorter than 16 is ideal for jet fuel (Olmstead et al., 2013; Williams and Laurens, 2010).

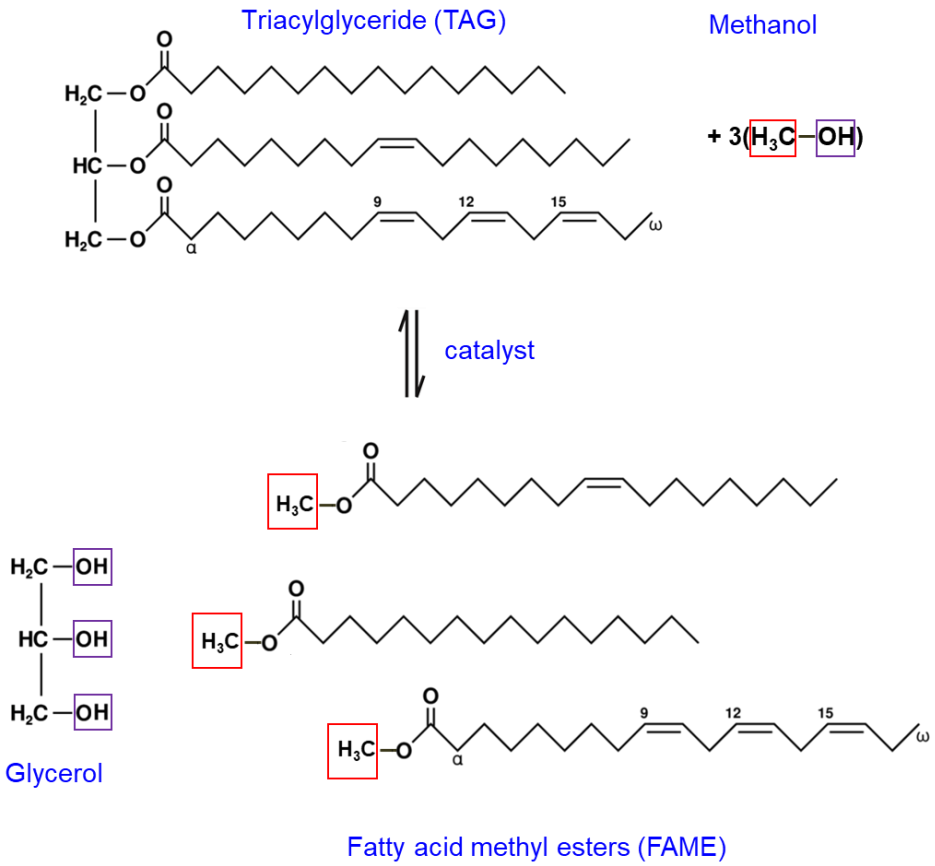


Figure 1-5 Transesterification of TAG to FAME. The FAMEs are used for biodiesel. The alcohol may also contain a different alkyl group. The catalyst is often hydrochloric acid or potassium hydroxide. Not only TAGs may be esterified, this reaction is also used to analyse algal lipids by their FAME profile across different lipid classes.

1.2.2.3 Economic and downstream considerations of algal biofuels

There was a peak of interest in microbial and microalgal sources of oil in the 70s and 80s but subsequently the drop in crude oil prices meant algal biofuel production was economically unviable by the 90s (Williams and Laurens, 2010). Despite there being multiple commercial biofuel projects around the world (reviewed in (Su et al., 2017)) it is still the consensus that currently to industrialise algae biofuel needs more basic research and development as evident from the issues discussed below.

One problem is the intrinsic limit on operational density in phototrophic systems due to light penetration and problems with culture pH control at high cell

density. Additionally, at high latitudes the lack of sun limits growth (Paranjape et al., 2016). It has been shown that theoretical yield of microalgae biofuel does not match that of the actual, typically being about three to eight times less than predicted (Show et al., 2017). Mixotrophic and heterotrophic growth strategies can allow cultivations at higher cell densities, but one must consider the cost of input of the organic carbon source, and also this detracts from the (often overly-romanticised) vision of biofuel using only sunlight, basic inorganic nutrients, water and carbon dioxide via photosynthesis. There has been research on using low-cost and waste carbon sources such as glycerol; however, the ability of algal strains to utilise these sources varies greatly (Paranjape et al., 2016) and therefore puts another parameter onto the 'ideal' production strain which must be screened for or engineered. Also, rich carbon sources can be a risk of contamination in commonly used open-pond systems (Chew et al., 2018). The use of bespoke bioreactors suitable for high density growth has also been investigated but the overhead costs of these are high and can require additional energy and cost inputs for lights and temperature control. Some bioreactors use alternative physical technologies such as mirrors and phosphorescence to attempt to overcome light limitation (Raha et al., 2018).

Although recognized as promising, the hype surrounding the research of biofuels from microalgae has also come under criticism. For example, although there are individual algal species with required characteristics for industry, such as high salt tolerance, high productivities, high hydrocarbon content, containing high-value products and those suitable for harvesting techniques such as auto flocculation (**Figure 1-6**), they occur in separate species rather than in the same organism and even genetically engineered strains are yet to reach all these requirements, therefore it remains economically unviable (Lim and Schenk, 2017; Klein-Marcuschamer et al., 2013). The acknowledgement of the need for basic research has been met by government input for funding when there is a lack of competitive market. It is recognised that research into algal biofuels is still important, which is evident from funding for projects around the world including governmental investments in America, the EU, and discussion at the COP21 United Nations Climate Change

Conference (Lammers et al., 2017; NAABB, 2014; Pires, 2017; Su et al., 2017; Unkefer et al., 2017).

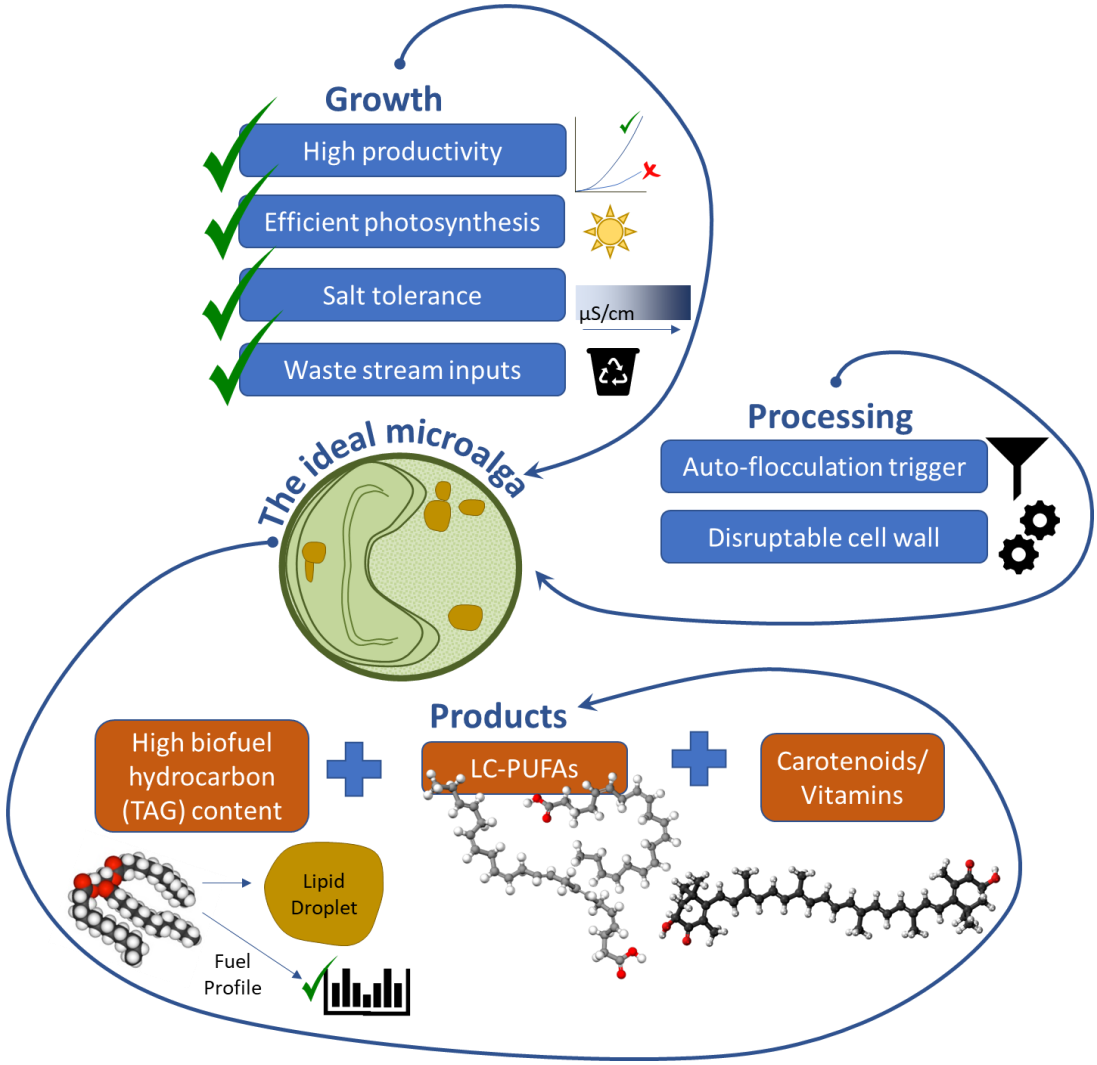


Figure 1-6 Characteristics of an ideal microalga for production of bioproducts. At present algal strains do not meet all of these requirements.

It is generally accepted that a combined low and high-value product 'biorefinery' approach with waste stream process inputs and carbon dioxide capture is necessary in order to be competitive, and that the production of high-value chemicals represents a lifeline for 3rd generation biofuels (Borowitzka, 2013; Chen et al., 2015; Moreno-Garcia et al., 2017; Zhu, 2015). Therefore, some research and commercial interests have shifted and diversified towards high-value products from microalgae, in fact this was essential for some biofuel companies to remain viable (Su et al., 2017). The screening and genetic engineering continues for strains containing

multiple desirable traits and so-called 'triple-producer' strains as described in **Figure 1-6** (Schüler et al., 2017). In new research the need for 'life-cycle' assessment is being increasingly realised (Kern et al., 2017).

1.2.3 Nutritional and High-value compounds

The use of microalgae as a food source is not a new concept and there are historical accounts of the consumption of Spirulina cyanobacteria Spirulina (*Arthrospira platensis* and *A. maxima*) dating back to Aztec times in Mexico, and it is part of normal daily diet for Chad in Africa (Becker, 2013; García et al., 2017). The cyanobacterium *Nostoc* is also widely consumed as food around the world, as are some filamentous green algae (García et al., 2017). Macroalgae are perhaps more widely consumed, such as the red seaweed *Pyropia* or 'nori', most popular in Japan (Wells et al., 2017). In western supermarkets, microalgae consumption is less mainstream and tends to only be sold as a food supplement or health food, typically being raw microalgal biomass such as *Chlorella* sp. or *Spirulina* sp. However, the high protein content of microalgae led to interest in the 1950s in microalgae being a possible strategy to cope with increasing global food demand (Fisher and Burlew, 1953). Japan were the first country to have industrial scale production of *Chlorella* sp. for human consumption (Vigani et al., 2015) and around 90 percent of algal cultivation is reported to be located in Asia (Sathasivam et al., 2017).

The composition of microalgal biomass differs greatly between species but as an example, commercially available *Chlorella* sp. biomass can contain up to 70% dry weight in protein including all nine essential amino acids required in human diets (Wells et al., 2017). The *Chlorella* sp. is typically used as a protein powder supplement similar to whey protein powders which are sold at typically 35-80 % protein (Henchion et al., 2017). However, in addition to contributing to normal nutrition, microalgae such as *Chlorella vulgaris* are reported to have additional health benefits such as anti-cancer and immune-modulating properties (Safi et al., 2014).

The production of algal biomass is not just for human nutrition. The use of algae for animal feed is common, including the aquaculture industries of which 50 % of globally consumed fish is supplied currently (Henchion et al., 2017). The market

value of the algal biomass is reportedly 100 €/kg for human consumption compared to 5-20 €/kg for agriculture and aquaculture feed or 0.40 €/kg for biofuel (Sathasivam et al., 2017) (**Figure 1-7**). Algal astaxanthin has a market value of up to 3000 USD/kg (D'Alessandro and Antoniosi Filho, 2016). The market value of various high value compounds and the economics of algal production is reviewed in other reports along with the state of research and development in different microalgal species (D'Alessandro and Antoniosi Filho, 2016; Hallmann, 2007; Leu and Boussiba, 2014; Martins et al., 2013; Stanley et al., 2013).

Figure 1-7 The market value of various products from microalgae and macroalgae. Reproduced from the Algal Bioenergy Special Interest Group Report: A UK Roadmap for Algal Biotechnologies (Stanley et al., 2013).

Other specific purified chemicals are sold as supplements with health or medical benefits such as the 'cosmaceuticals' e.g. squalene in skin care products or 'nutraceuticals' e.g. beta-carotene from *Dunaliella salina*, astaxanthin from *Haematococcus pluvialis*, and long-chain polyunsaturated fatty-acids (LC-PUFAs), which are sold as food supplements (Borowitzka, 2013; Leu and Boussiba, 2014). One problem is that some of these so-called 'neutraceuticals' are lacking regulatory

frameworks because they fall in a grey area between a food supplement and a pharmaceutical, and claimed health benefits may not be properly substantiated by evidence or screened for safety (Santini et al., 2018). The development and sale of these products by poorly-regulated companies can undermine the ethics of algal research.

1.2.3.1 Long-chain polyunsaturated fatty acids (LC-PUFAs)

Of the nutritional compounds, LC-PUFAs (**Figure 1-8**) are of key importance to human diet, with roles in development, cardiovascular disease prevention, anti-inflammatory responses, and brain function including treatment of schizophrenia (Amminger et al., 2015; Bellou et al., 2016; Robertson et al., 2015; Ryckebosch et al., 2014, 2012). The omega-3 fatty-acids EPA (eicosapentaenoic acid; $20{:}5\Delta^{5,8,11,14,17}$, n-3) and DHA (docosahexaenoic acid; $22{:}6\Delta^{4,7,10,13,16,19}$, n-3) are the two main essential LC-PUFAs and are the focus of this work (**Figure 1-8**). They are called 'essential' because they are unable to be synthesised to adequate amounts in both higher plants and animals, and the PUFA ALA (C18:3n3), the precursor to all other n3 and n6 PUFA, is unable to be synthesised at all (Pereira et al., 2012). Those animals such as fish which do contain them, obtain them through their diet from the primary producers, i.e. microalgae.

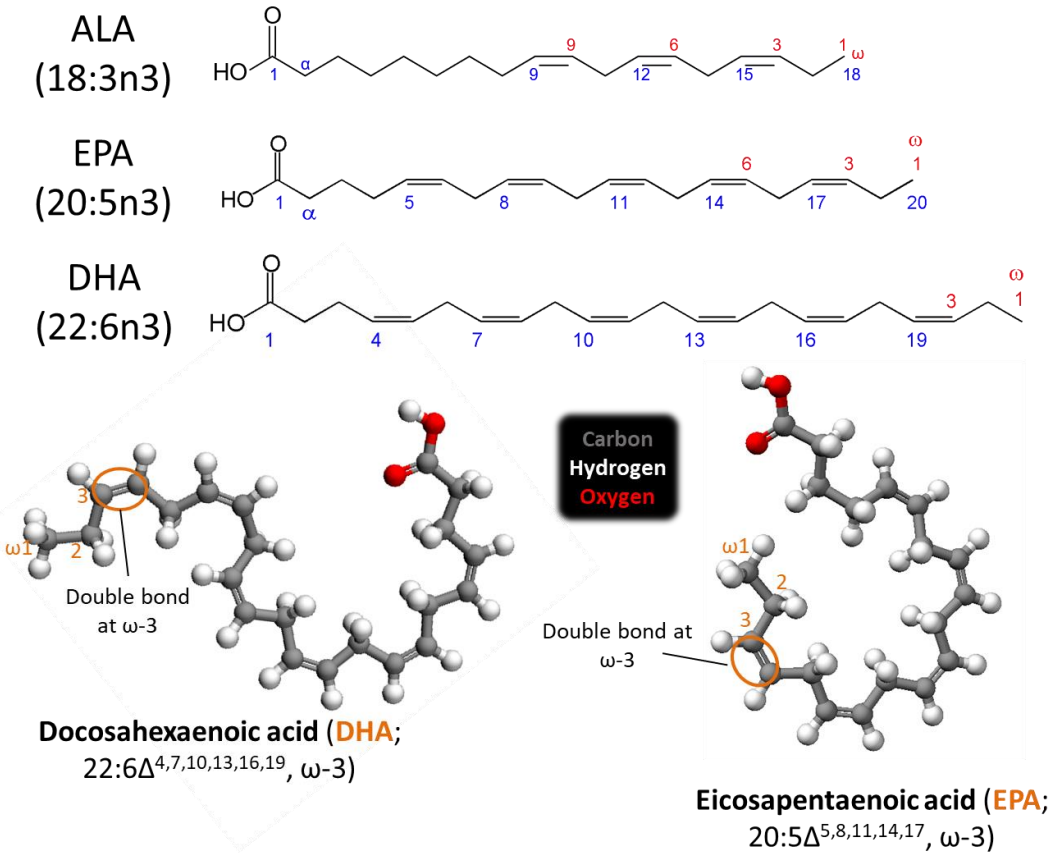


Figure 1-8 Structures of nutritional LC-PUFA omega-3 fatty acids. The carbon chains are labelled from the omega or "n" end. ALA is alpha linoleic acid; 18:3Δ^{9,12,15}, ω-3. The structures of EPA and DHA demonstrate the differences in physical characteristics of carbon chains with differing chain lengths and desaturations. EPA and DHA Structures were downloaded from PubChem (NCBI) and 3D image was generated in Accelrys Disccovery Studio 4.1.

Traditional dietary sources of these compounds are oily fish including herring, mackerel, sardine and salmon (Ruiz-Lopez et al., 2014b). However, wild stocks and fish farms will not be able to cope with increased global demand. Furthermore, there are concerns regarding accumulation of environmental pollutants (e.g. mercury) in fish, problems with perceived offensive smell/taste, and unsuitability of fish supplements for individuals following a vegetarian diet. So alternative sources are needed (Brunner et al., 2009; Ruiz-Lopez et al., 2014b; Ryckebosch et al., 2014; Tang et al., 2016a). Some algae contain naturally high percentages of EPA or DHA within their lipids, such as the diatom *Phaeodactylum tricornitum* (Lang et al., 2011), the

marine alga *Nannochloropsis gaditana* (Bellou and Aggelis, 2012; Slocombe et al., 2015), and freshwater alga *Monodus subterraneus* (Khozin-Goldberg et al., 2002).

Microalgae are ultimately the primary producers of EPA and DHA. The levels of EPA and DHA in high producers range from 21-45 % which is comparable to many other microbial sources including bacteria and fungi (see Adarme-Vega et al., 2012 for a review). The screening and characterisation of culture collections is a valuable way of identifying promising production strains and was successfully used to identify a salt-tolerant strain of *Chlorella* sp. with high fat content containing a beneficial dietary ratio of ALA compared to LA (linoleic acid) (Slocombe et al., 2013). LA is an omega-6 fatty acid whereas ALA is omega-3. It has been shown that western diets have too much omega-6 compared to omega-3 (Pereira et al., 2012).

Interestingly, 70% of fish oil is used in aquafeeds and there is an acknowledged limit in the supply of EPA and DHA from aquaculture sources which leaves the market open for other sources such as microalgal biomass (Chauton et al., 2015). Microalgae are already an established alternative source of DHA in infant formula (Adarme-Vega et al., 2012). There have been successful efforts to modify the metabolism of the EPA-producing microalga *P. tricornutum* to additionally produce DHA by expressing heterologous genes in the omega 3 pathway. This resulted in an 8-fold increase in DHA content from trace levels (Hamilton et al., 2014). Other sources of EPA and DHA include genetically engineered oilseed crops such as *Camelina sativa* (Ruiz-Lopez et al., 2014b, 2014a), though these have a disadvantage in reduced biomass productivity compared to algal systems of production. The market for EPA and DHA is reportedly 300 million and 1.5 billion USD respectively, so they are an attractive product from microalgae.

1.3 Fatty-acid metabolism pathways and function in microalgae

There has been increasing interest in microalgal lipid metabolism over the past decade, with a growing number of publications (Liu and Benning, 2013). In order to manipulate the lipid profile towards such preferable products such as TAGs or a

specific LC-PUFA, it is important to examine the lipid metabolism in microalgae, and more specifically green microalgae in relation to *C. sorokiniana*.

The lipid pathways in microalgae are not fully elucidated although bottlenecks and key enzymes have been identified by transcriptomic studies (Fan et al., 2015; Klok et al., 2014; Mühlroth et al., 2013). A lot of initial insight was gained through comparisons of microalgae with plant lipid metabolism, however, that itself is not fully elucidated either and these insights are not necessarily applicable to microalgae with many differences found between plant and algal pathways (Liu and Benning, 2013; Merchant et al., 2012). Also, within microalgae, lipid metabolism itself is diverse. For example, this is highlighted in the model alga *Chlamydomonas* which has no phosphaditylcholine (a membrane phospholipid), although this is present in other microalgae, and in plants it has a crucial role in the transfer of *de novo* fatty-acids from the plastid to the acyl CoA pool in the cytoplasm. In the case of *Chlamydomonas* betaine lipids such as DGTS (1,2-diacylglyceryl-3-O-4'-(N,N,N-trimethyl)-homoserine) fulfil this role (Liu and Benning, 2013).

Ultimately a simplified general scheme can be provided as an overview which has been reviewed many times within several different contexts (**Table 1-1**). Using these references, a general scheme for algal lipid metabolism and the factors which affect it is discussed below (**Figure 1-9**).

Context	References
Oleaginous microbes	(Bellou et al., 2016; Garay et al., 2014)
Microalgae in general	(Bellou et al., 2014; Khozin-Goldberg, 2016; Schüler et al., 2017)
Using the model *Chlamydomonas reinhardtii*	Chlamydomonas (Boyle and Morgan, 2009; Gargouri et al., 2015; Johnson and Alric, 2013; Li-Beisson et al., 2015; Liu and Benning, 2013; Merchant et al., 2012)
In *Chlorella* sp.	(Fan et al., 2015; Wu et al., 2015)
Within the context of biofuels	(Hu et al., 2008; Williams and Laurens, 2010)
Focusing on TAGs	(De Bhowmick et al., 2015; de Jaeger et al., 2014; Lenka et al., 2016; Merchant et al., 2012; Schüler et al., 2017; Zienkiewicz et al., 2016)
Focusing on high-value products	(Bellou et al., 2016; D'Alessandro and Antoniosi Filho, 2016; Garay et al., 2014; Klok et al., 2014; Martins et al., 2013; Minhas et al., 2016; Schüler et al., 2017)
Whole cell carbon metabolism and metabolic networks	(Boyle and Morgan, 2009; de Jaeger et al., 2014; Fan et al., 2015; Gargouri et al., 2015; Johnson and Alric, 2013; McKie-Krisberg et al., 2018; Wu et al., 2015; Xie et al., 2015; Zhan et al., 2017)
Modelling approaches	(Lenka et al., 2016)
Molecular basis and potential target genes for over/under-expression	(Cecchin et al., 2018; Li-Beisson et al., 2015; Merchant et al., 2012; Perez-Garcia et al., 2011; Zienkiewicz et al., 2016)

Table 1-1 Reviews covering lipid metabolism of microalgae in various contexts. Note some references are repeated.

Fatty acids are an important part of the cell physiology of algae. After their *de novo* synthesis in the chloroplast by the type II FAS which yields C16:0 (Palmitic) and C18:0 (Stearic) acyl chains attached to acyl carrier protein (ACP), they are destined for multiple fates (**Figure 1-9**). Firstly, it should be noted that the C18:0-ACP may undergo an initial desaturation reaction in the chloroplast to yield C18:1n9-ACP (oleic) and is therefore often classed as a third species yielded from the FAS II cycle (**Figure 1-9**). The fates of these three species may then be: incorporation into phospholipids (also called glycerophospholipids), which form part of the cell membranes, specific glycoglycerolipids such as MDGD (monogalactosyl-diacylglycerol), DGDG (digalactosyl-diacylglycerol) and SQDG (sulfoquinovosyl-diacylglycerol) which form the thylakoid membranes, and incorporation into storage glycerolipid triacylglycerol (TAG) which can accumulate as lipid droplets in both the cytosol and the plastid. Some fatty acids undergo additional elongation and

desaturation reactions to yield long-chain poly-unsaturated fatty acids (LC-PUFA) such as EPA and DHA, which are then incorporated into the different lipid classes (**Figure 1-10**). Some of these fates, such as LC-PUFA synthesis, rely on export of fatty-acid acyl chains out of the chloroplast where they are then attached to Coenzyme-A (CoA) to form the Acyl-CoA pool (**Figure 1-9**). There is a large diversity of fatty-acid chain length across different microalgal species and not all microalgae synthesise LC-PUFAs naturally. For example, *Chlorella sorokiniana* UTEX1230 only naturally accumulates fatty acid up to ALA (C18:3n3) and therefore requires expression of heterologous enzymes to complete the omega-3 pathway (**Figure 1-10**).

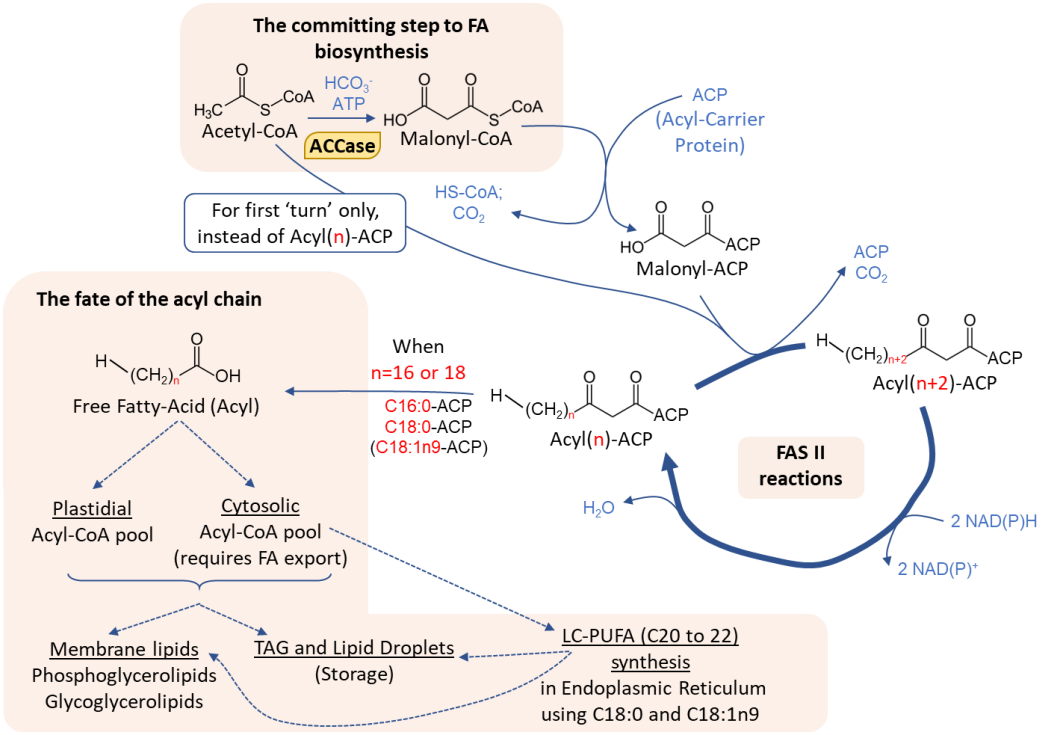


Figure 1-9 Summary of fatty-acid metabolism in microalgae. *De novo* synthesis of the fatty-acid chain occurs in the chloroplast. Acetyl-CoA precursors fate is determined for fatty-acid biosynthesis once the reaction to malonyl-CoA is catalysed by Acetyl-CoA carboxylase (ACCase). On the first initiating step of the Fatty Acid Synthase Type II (FAS II) cycle, Acetyl-CoA and Malonyl-ACP (ACP: acyl-carrier protein) are the substrates to form the first acyl-ACP. After this, the FAS II reactions sequentially add on two carbons to the growing acyl-ACP chain for each turn of the cycle until the acyl chain reaches 16 or 18 carbons, at which point it is released from ACP to become free fatty acid (FA). The FAS cycle is overall a condensation and requires reducing power as NAD(P)H. The main possible fates of the nascent FA are shown. Although not included in the diagram, there is evidence of transfer from the phospholipids and TAGs via enzymes such as PDAT (phospholipid:diacylglycerol acyltransferase). Figure drawn using the following references as a guide (Fan et al., 2015; Garay et al., 2014; Khozin-Goldberg, 2016; Li-Beisson et al., 2015; Schüler et al., 2017).

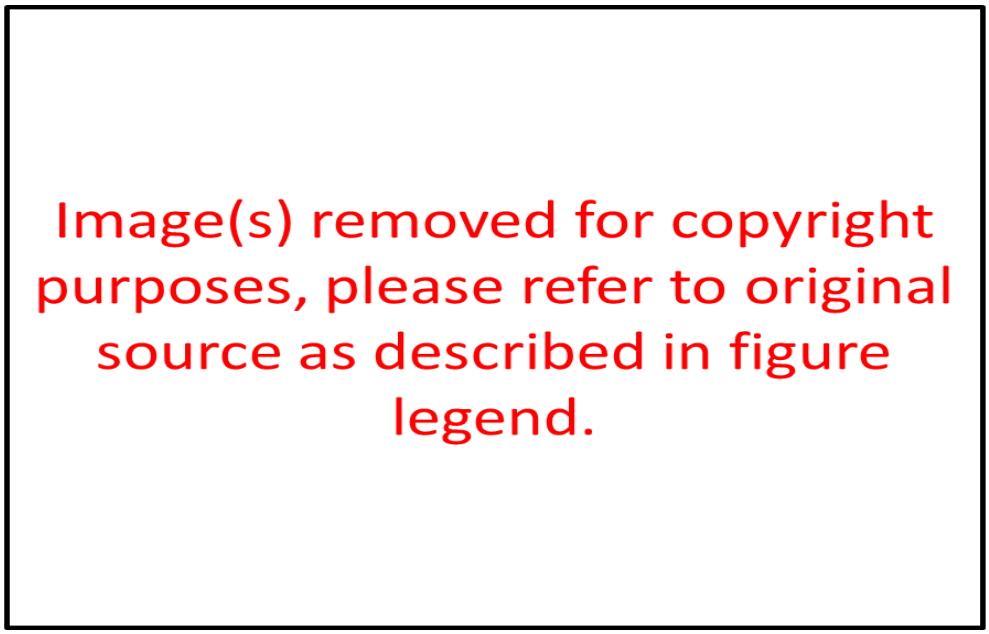


Figure 1-10 LC-PUFA synthesis pathways in microalgae. Using the mono-unsaturated fatty acid oleic acid (OA, C18:1n9), which is one of the outputs from fatty-acid synthesis in the chloroplast, the carbon chain undergoes sequential elongation and desaturation reactions following either an omega-3 or omega-6 pathway. This takes place in the endoplasmic reticulum and therefore the OA must be exported out of the chloroplast to join the cytosolic acyl-CoA pool before LC-PUFAs can be synthesised. This figure is modified from Bellou *et al.* (Bellou et al., 2016).

Examples of the types of lipid and different classes are summarised in **Table 1-2**. The multiple types of membrane lipid have different functions in terms of spatial packing. For example, PA (phosphatidic acid) and PE (phosphatidyl-ethanolamine) have a large tail area compared to their head, and therefore can induce concave curvature in a membrane whereas PI (phosphatidyl-inositol) has a small tail area compared to the head and therefore induces convex curvature, and PC (phosphatidyl-choline) is cylindrical (Garay et al., 2014). This is the same for the photosynthetic glycolipids, where MGDG tends to be conical but DGDG and SQDG are cylindrical (Da Costa et al., 2016). PG is the only phospholipid in thylakoid membranes but it is also found in extraplastidic membranes, and is the only type that contains the acyl species 16:1Δ3-trans at the sn-2 which is important for photosynthesis at low temperatures (Khozin-Goldberg, 2016). In contrast to membrane lipids, TAGs are used by the cell as a form of carbon storage and require reducing power in the form

of NAD(P)H during their synthesis (**Figure 1-9**). In TAG biosynthesis the terminal acyltransferases are seen as the rate-limiting step (Khozin-Goldberg, 2016).

The patterning of the acyl chains in the various positions in TAG and DAG derivatives can also give information about their pathway of synthesis. For example chains of 16 carbons at the sn-2 position suggest synthesis by the 'prokaryotic' chloroplast pathway and chains of 18 carbons at the sn-2 position are characteristic of the 'eukaryotic' pathway in the endoplasmic reticulum (Khozin-Goldberg, 2016; Li-Beisson et al., 2015; Liu and Benning, 2013). Acyl chains may be transferred between these different pools, for example polar membrane or thylakoid lipids from various locations can be sources for neutral lipids by cell remodelling using enzymes such as PDAT (phospholipid:diacylglycerol acyltransferase) (Garay et al., 2014; Khozin-Goldberg, 2016). Alternatively, neutral lipids may be synthesised *de novo* and key metabolic intermediates can pool in strategic locations (Garay et al., 2014).

It can be seen that there are many metabolic elements involved in fatty-acid synthesis (**Figure 1-9**). Ultimately, the fate of the fatty acids relies on the availability of the acyl-CoA pool and its location, and the availability of the other key substrates such as G3P (glycerol-3-phosphate) and acetyl-CoA. However, these elements are an important part of other cellular metabolism too and availability depends on the balance between other processes such as the Calvin cycle and glucogenesis. Therefore, consideration of the whole cell metabolism is required to understand the mechanisms properly, but this is beyond the scope of this work. As seen in **Figure 1-9**, fatty acid synthesis occurs in the plastid by the complex Fatty Acid Synthase. The inputs for this are Malonyl CoA, from acetyl CoA with synthesis also requiring reducing power in the form of NAD(P)H. The enzymes are omitted from **Figure 1-9** apart from the one catalysing the committing step the FA biosynthesis, ACCase, which has been a target for genetic engineering approaches for increasing lipid content in microalgae (Li-Beisson et al., 2015). In terms of how neutral lipid accumulation as energy storage fits into the rest of the cell metabolism, it has been shown that the cell needs more reducing power (NAD(P)H) and ATP than other storage compounds such starch and other polysaccharides derived from glucose (Garay et al., 2014; MacDougall et al., 2011; McKie-Krisberg et al., 2018; Williams and Laurens, 2010).

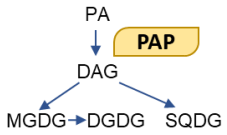
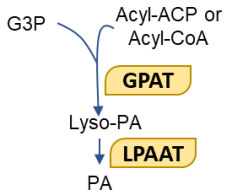
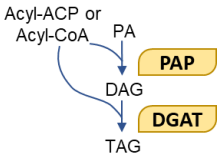
Lipid name	Type	Function	Key precursors and enzymes involved
MGDG and DGDG (mono- or di-galactosyl-diacyl-glycerol)	Glyco-glycero-lipid (polar)	Thylakoid photosynthetic membranes	 Synthesised in chloroplast
SQDG (sulfo-quinovosyl-diacyl-glycerol)	Sulfo-glycero-lipid (polar)		
PG (phosphatidyl-glycerol)	Phospho-glycero-lipid (polar)	Cell membranes (especially phospholipid bilayer), cell signalling. PG occurs in both plastidial and non-plastidial membranes.	Ultimately PA, which yields DAG and CDP-DAG precursors. Reaction details not shown, but are reviewed in (Khozin-Goldberg, 2016). Synthesised outside the chloroplast except PG can be both.
PE (phosphatidyl-ethanolamine)			
PC* (phosphadityl-choline)			
PI (phosphatidyl-inositol)			
PA (phospatidic acid)		Minor constituent of membranes, important intermediate to all glycerolipids	
TAG (triacyl-glycerol)	Glycero-lipid (neutral)	Storage lipids. Desirable form for biofuel conversion	 Synthesised in Endoplasmic Reticulum ("Kennedy Pathway") and chloroplast
DGTS	Betaine lipid	Functional equivalent of PC in some microalgae. Other betaine lipids exist.	DAG (details not shown, refer to (Khozin-Goldberg, 2016)

Table 1-2 Common lipid products present in microalgae. DGTS: 1,2-diacylglyceryl-O-4-(N,N,N-trimethyl)homoserine. PAP: phosphatidate phosphatase. GPAT: glycerol-3-phosphate acyltransferase. LPAAT: lyso-phosphatidic acid acyltransferase. DGAT: diacylglycerol acyltransferase. Lyso-PA: lyso-phosphatidic acid.

1.4 Strategies for enhancing LC-PUFA and lipid content in microalgae

Although one strategy for improving microalgal production strains is to bio-prospect for new strains and screen for a desirable lipid profile, there is much research directed towards enhancing desirable parameters of currently characterised strains. There are different aims in this: firstly, to increase the total neutral lipid content for biofuel purposes, and secondly, to obtain an ideal fatty acid chain profile. Depending on the purpose of the production strain, these requirements will differ. As discussed previously, the neutral lipids such as TAGs, which accumulate as oil bodies, are preferred for biofuels because the phosphorus and sulphur present in other lipid classes is inhibitory to the transesterification process used to make biodiesel from extracted lipids (which therefore affects the total yields), and also sulphur is a contaminant in engines (Li et al., 2008; MacDougall et al., 2011; Williams and Laurens, 2010). In terms of lipid profile, saturated or mono-unsaturated fatty acids are generally preferred with a typical chain length of 18 carbons, though shorter chains are ideal for jet fuel (Olmstead et al., 2013; Williams and Laurens, 2010). Ideally, strains would exhibit multiple desirable characteristics to allow for harvesting of several useful products.

The screening of existing microalgal culture collections highlighted some interesting observations including strains which had anomalous profiles for their taxa. For example, ARA (arachidonic acid) was found in exceptionally high amounts (up to 73.8 % or 120 ug/mg dry weight) in the chlorophyte *Palmodictyon varium* SAG 3.92 (Lang et al., 2011). However, the authors also note that predicting PUFA contents at species and genera levels is unreliable and hence during bioprospecting each strain should be examined individually (Lang et al., 2011).

Multiple strategies for altering TAG content and the fatty acid profile of microalgae have been used by the research community. These are discussed below and include manipulation of growth conditions and genetic or metabolic engineering to up or downregulate key enzymes in the lipid biosynthetic pathways. Strategies for increasing LC-PUFA include screening for new species that are naturally high in these products (Slocombe et al., 2015), assaying and determining optimal growth

conditions for lipid accumulation (Hallenbeck et al., 2015a, 2015b, 2015c), or genetic engineering (Hamilton et al., 2014; Levering et al., 2015).

1.4.1 The importance of productivity

Some species of microalgae are naturally oleaginous, accumulating lipid contents of up to 75 % of cell dry weight (for example *Botryococcus* sp.) but more commonly up to 50 % (*Chlorella* sp., *Nannochloropsis* sp. and *P. tricornutum*) (De Bhowmick et al., 2015). However, there is large variation between reports that probably reflects either strain or species variations, or various cultivation conditions (D'Alessandro and Antoniosi Filho, 2016). Additionally, the lipid content of cells must be balanced with the productivity and growth rate in terms of biomass per litre of culture per day to be economically viable (Draaisma et al., 2013). For example, *Botryococcus* sp. productivity in terms of biomass was only 0.02 $gL^{-1}d^{-1}$ whereas for *P. tricornutum* it was 0.3-1.4 $gL^{-1}d^{-1}$, showing that although *P. tricornutum* has a lower overall lipid content it could be a better strain to use for commercial production (De Bhowmick et al., 2015). *C. sorokiniana* has shown to have productivities of 1.47 to 12.2 $gL^{-1}d^{-1}$ and although in these cases only had lipid of 19-22 % dry cell weight, the productivity compensates for this comparatively lower lipid content (De Bhowmick et al., 2015). There is also an economical balance in terms of processing cost, because higher volumes of biomass are more costly to process by procedures such as centrifugation. Though for a high-value product, such as the nutritional omega-3 fatty acids EPA and DHA, lower productivities and amounts are tolerated due to the high market value of the final product.

The productivity of a culture depends on the input of nutrients and its physical growth conditions. Highest productivities are gained in intensive cultures in photobioreactors, but these have higher setup costs than other systems such as open ponds. However, a benefit of using bioreactors is that more precise control can be achieved, and they do not rely on the external environment. Therefore, issues such as the weather are not a problem which means they can be grown all year round. They are also less likely to get contaminated than open pond systems, which can be a significant problem for outdoor cultivation.

There is a large diversity in the species, and indeed strains, that have been characterised for lipid production (Bharathiraja et al., 2017; D'Alessandro and Antoniosi Filho, 2016; Schüler et al., 2017). One issue is that it is difficult to compare productivities across these isolated reports because of the many differences in growth conditions ranging from base media to abiotic factors. Additionally, productivity may not be directly reported, as was found to be the case in a survey of published data on *Nannochloropsis* sp. (Xu and Boeing, 2014). The authors of the survey also argue that a more valuable measure to be reported would be maximum sustainable yield, which is better aligned with fishery management and wildlife management sectors to which the oil product is a competitor (Xu and Boeing, 2014).

1.4.2 Abiotic growth conditions and nutrient stresses

The factors affecting microalgae lipid accumulation and profile have been extensively surveyed and reviewed many times (Chen et al., 2017; De Bhowmick et al., 2015; Minhas et al., 2016; Sayanova et al., 2017; Zienkiewicz et al., 2016). These can be divided into abiotic conditions such as light, temperature, salinity, pH and nutrient stresses such as deprivation, addition of chemicals and modification of growth mode to mixotrophy or heterotrophy (supply of external carbon sources). Many studies investigate a couple of parameters on growth. Although this is valuable, it would be beneficial to employ a strategy like that was first employed by Evens *et al* and subsequently Wolf *et al*., Hallenbeck *et al.* and others (Evens and Niedz, 2010; Hallenbeck et al., 2015a, 2015b, 2015c; Kanaga et al., 2016; Wolf et al., 2015), who conducted multi-parameter high-throughput screens to optimize nutrient requirements. They highlight that multi-parameter response optimisation gives better performance than optimisation of single parameter responses and that optimal values differ when considered in combination, and that there are strain-specific differences even between the same species. Additionally, Kanaga *et al.* report that their photoautotrophic growth rates obtained as a result of their optimisations are amongst the highest reported (Kanaga et al., 2016).

Perhaps the most well-known strategy for inducing lipid accumulation, especially in neutral lipids such as TAGs, is to place the cells under the nutrient stress

of depleted nitrogen (Chen and Johns, 1991; Draaisma et al., 2013; Klok et al., 2014; Li et al., 2016). A disadvantage with this approach is the impeded growth rate that results from the stress, since cells become unable to divide (the premise of redirecting their metabolism to storage compounds). Although, one study using *Chlorella sorokiniana* reported an unimpeded growth rate under nitrogen stress for up to two weeks (Negi et al., 2015), this is not common. This stress response is universal across microalgal diversity; for example in both *Chlorella* sp. and *Nannochloropsis* sp. which are in different evolutionary groups (**Figure 1-1**), depleted nitrogen increases the proportion of neutral lipid, reduces glyco- and phospho-lipid, and increases the total lipid per biomass, showing that although some of this is cell remodelling, it is also extra synthesis (Olmstead et al., 2013). However, it was recently shown in a thorough multivariate screen that unexpectedly there was little correlation between nitrogen content and lipid accumulation (Hallenbeck et al., 2015a). This shows that even well-established observations may have hidden complexities due to confounding effects of ionic context (Evens and Niedz, 2010). Other nutrient stresses resulting in neutral lipid increases include deprivation of phosphorus (Wu et al., 2013) and sulphur (La Russa et al., 2012). These effects can also act in synergy, as shown with nitrogen and phosphorus for *Chlorella minustissima* (Arora et al., 2015) and *Chlorella protothecoides* (Yuqin Li et al., 2014).

Carbon supplementation can alter the trophic mode of the microalga from phototrophy to mixotrophy or heterotrophy, when the microalga obtains energy from uptake of the organic molecules instead of, or in combination with, fixation of atmospheric carbon dioxide. Different trophic modes have been shown to induce a significant effect on the lipid content of microalgae. Heterotrophy is typically associated with an increase in lipid (Miao and Wu, 2006; YuQin et al., 2015). For example, in *C. sorokiniana* UTEX 1230, heterotrophic cultivation yielded twice as much TAGs compared to phototrophy (Rosenberg et al., 2014). Another example is *C. protothecoides* which was reported to have four times the amount of total lipids in heterotrophy compared to phototrophy (Xiong et al., 2008). These changes in lipid composition are a result of a metabolic shift in the carbon flux of the cell from carbohydrate production to lipids (Ren et al., 2016). It can also alter the fatty acid

profile; in the case of *C. prototheco*ides supplementation with glucose resulted in increased saturated fatty acids in the profile (Ren et al., 2016). For *C. vulgaris*, supplementation of 10 g/l glucose resulted in the same final cell density as the control but a faster growth rate and nearly double the dry weight of cells. However, the PUFA content was lower (Kirchner et al., 2016), so this strategy may be more relevant for a biofuel application than for increasing high-value omega-3 lipids.

Light intensity can affect the lipid content of the cells because light can induce a metabolic shift in carbon metabolism (Fan et al., 2015). The effects vary between species and are reviewed by Chen *et al* (Chen et al., 2017). For example in *C. sorokiniana*, a mixotrophic cultivation strategy resulted in a three-fold increase in lipids relative to the absence of light in true heterotrophy (T. Li et al., 2014). However, for *C. zofingiensis*, supplementation of glucose-fed cultures with light reduced the lipid content and increased starch instead (T. Chen et al., 2015).

Other strategies include culturing with growth promoting bacteria by immobilising on alginate beads (Leyva et al., 2014). Chemical inducers of lipid accumulation have also been investigated: for example Brefeldin A, a drug known to affect the Golgi apparatus (Fujiwara et al., 1988), has been found to be effective in *C. reinhardtii* and to a lesser extent, *C. sorokiniana* UTEX 1230 (Wase et al., 2015). Overall, the effect of growth conditions on lipid metabolism in microalgae is a highly complex multifactor process hence there can be contradictive responses reported in the literature. This is an inherent part of their physiological response to, and survival in, a changing environment.

1.4.3 Genetic engineering

Although there has been success with different culturing conditions and supplements in changing lipid profiles and increasing TAGs, ultimately an understanding of the lipid metabolism pathways and genes involved would allow for increased control through metabolic engineering by upregulating or downregulating certain genes involved in lipid metabolism to direct the cell towards these products. Genetic engineering can also be used to introduce entirely new pathways into an organism to make non-native or non-natural fatty-acids.

Genes which have shown some effect on lipid content in various studies include diglyceride acyltransferase (DGAT) which is involved in the final TAG-committing step of the Kennedy pathway (see **Table 1-2**), and acyl-ACP esterase (AAE) which is involved in hydrolysis of the mature fatty-acyl chains to create a pool of free fatty acids (Klok et al., 2014). For DGAT, this enzyme has been identified as a rate limiting step in TAG biosynthesis in plants (Lung and Weselake, 2006). There are, in fact multiple DGATs in microalgae but the specific function of each one has not been fully elucidated. When characterised in the oleaginous microalga *C. vulgaris*, DGAT1 was found to be expressed constitutively at a high level, unlike in *C. reinhardtii*. The authors suggest that the higher DGAT1 expression could explain the relatively high oil content (40 % by dry weight) of *C. vulgaris* compared to *C. reinhardtii* (Kirchner et al., 2016).

For increasing TAG, one metabolic engineering strategy is to eliminate competing pathways such as starch biosynthesis. This was successfully employed for the microalga *C. reinhardtii* to increase TAG content, but the level of TAG still did not reach that of some naturally oleaginous microalgae (de Jaeger et al., 2014). A promising strategy would be to use this approach on an already oleaginous algae, as was done successfully for *Scenedesmus obliquus* without compromising total biomass or photosynthetic efficiency (Breuer et al., 2014). However, other evidence presented in *C. sorokiniana* demonstrated that this approach of simply reducing starch accumulation can be too simplistic, and that carbon partitioning is a more complex process (Vonlanthen et al., 2015). Characterisation of the transcriptional changes occurring in cells under nitrogen stress is another strategy to identify potential gene candidates for over-expression (Goncalves et al., 2016; Hernández-Torres et al., 2016; Vello et al., 2018; Zienkiewicz et al., 2016).

In terms of targeting the synthesis of LC-PUFAs, key steps for EPA and DHA involve lengthening the carbon chain via elongases and desaturation of the chain to form double bonds (**Figure 1-10**). Genes important for this have been identified and sequences are known in some green algae: for example delta-4 and delta-6 desaturases, and delta 5 elongase (Hamilton et al., 2014; Vaezi et al., 2013). Some of these have been used successfully to upregulate DHA in *P. tricornutum* (Hamilton et

al., 2014). Interestingly, for *Chlorella pyrenoidosa,* it seems that some lipid metabolism genes arose via horizontal gene transfer from bacteria (Fan et al., 2015). This suggests it could be feasible to widen the pool of potential genes to target to include those of oil-producing bacteria.

1.5 *Chlorella sorokiniana* UTEX 1230 as a strain of industrial interest

1.5.1 The *Chlorella* taxonomy and general characteristics

The now type-species of *Chlorella, Chlorella vulgaris* Beijerinck, was first formerly described in 1890 in an article in German by Marcus Beijerinck, who also re-classified some previously identified symbiotic "Zoochlorellae" to *Chlorella*, (Beijerinck, 1890; Shihira and Krauss, 1965). Although controversial at first, it is now the accepted designation (Proschold et al., 2011). These "Zoochlorellae" become especially relevant later in this thesis because free-living forms are susceptible to chloroviruses. Furhtermore, the symbiotic *Chlorella variabilis* NC64A (also known as CCAP 211/84) was the first *Chlorella* sp. to have a sequenced genome (Blanc et al., 2010). In the decades following Beijerinck's designation, approx. 1000 species of microalgae were assigned to this classification (Luo et al., 2010), with the definition of coccoid autosporic green cells covering a multitude of different species (Baudelet et al., 2017). Multiple subsequent, and sometimes competing (for example: (Kessler, 1976; Shihira and Krauss, 1965)), taxonomic revisions eventually led to "true" *Chlorella* sp. being identified, and the rest assigned different genera, based on molecular analysis (Baudelet et al., 2017; Champenois et al., 2015; Luo et al., 2010). *C. vulgaris* became historically important as a simple model for studying the physiological process in plants such as photosynthesis (Krienitz et al., 2015).

Some of the more well-known re-designated genera include *Parachlorella kessleri*, *Auxenochlorella pyrenoidosa and Auxenochlorella protothecoides* within the same class of Trebouxiophyceae, and *Chloromchloris zofingiensis* now in the sister class of Chlorophyceae (species name is the same, genus name for all of them was previously *Chlorella* sp.) (Baudelet et al., 2017; Krienitz et al., 2015). However, it is worth noting that these historical species are sometimes referred to colloquially as

within the general '*Chlorella* clade', (Luo et al., 2010). Of the now-resolved true *Chlorella* sp. there includes species of the typical green spherical cells: *C. vulgaris, C. lobophora* and *C. sorokiniana*. Additionally there are some species of the *Paramecium bursia* symbionts (Luo et al., 2010; Proschold et al., 2011) which are susceptible to infection by chlorovirus and affected the subsequent classifications of the viruses themselves, which are grouped according to the microalgae in which they infect. Finally, there are some *Chlorella* sp. which do not share the typically morphology, instead the may have mucilage envelopes and the ability to form colonies (Luo et al., 2010).

As one report only a few years ago states, the phylogeny of *Chlorella* sp. is a 'nightmare' and even at the molecular level can vary depending on the genetic marker used (18S rRNA or ITS2) (Heeg and Wolf, 2015), and it is likely that species revisions will continue to occur (Baudelet et al., 2017). Although it is valuable to have the correct taxonomic status, it is unfortunate that the many thorough physiological characterisations that were done in the early studies (for example Bonaventura and Myers, 1969; Sorokin, 1959; Sorokin and Krauss, 1959; Sorokin and Myers, 1953) are now difficult to link to their modern names, and so this information may be lost unless current authors are careful to acknowledge the old names.

Interestingly, these reclassifications mean their classification in food catalogues which are responsible for legislative control of the sale of these organisms, is not up to date and still refers to old names with the only three '*Chlorella*' appearing as *C. lutoviridis, C. vulgaris* and *C. pyrenoidosa* (Baudelet et al., 2017; Champenois et al., 2015; European Commission, 2015). Of these, *C. pyrenoidosa* is no longer a functioning taxonomic name. This is worth bearing in mind since so many industrial algal projects are based on making foodstuffs from microalgae and in primary research these downstream considerations are sometimes overlooked.

Other difficulties include that even at the species level there are obvious phenotypic differences between the multiple strains of *C. sorokiniana*. For example, there is *C. sorokiniana* UTEX 2805 which also tolerates growth conditions similar to UTEX1230 (de-Bashan et al., 2008), UTEX 2714 which is less thermotolerant (Bashan

et al., 2016) and a strain *C. sorokiniana* DOE 1412 whose maximum growth rate occurs around 37 °C (Lammers et al., 2017). Additionally, the common laboratory method of culture maintenance by propagation on nutrient agar encourages growth over multiple generations and genetic drift.

This strain variation is also seen within the fatty-acid profile. Interestingly, a lipid analysis of the entire SAG culture collection included five *C. sorokiniana* strains including 211-8k (equivalent to UTEX1230) and show high variation in the lipid profile as determined by FAME-GC or FAME-MS (Lang et al., 2011) (**Figure 1-11**A). In addition to standard fatty-acids, the authors also characterised other compounds, and surprisingly those present in *C. sorokiniana* include phytol, hexadecanamide (palmitamide) and 9Z-octadecenamide (oleamide). The latter two are primary fatty acid amides derived from palmitic acid (16:0) and oleic acid (18:1n9), respectively. Primary fatty acid amides are known to serve as signalling lipids (Ezzili et al., 2010; McKinney and Cravatt, 2005) and are not well characterised outside mammalian systems (Kim et al., 2010). Out of the five strains, *C. sorokiniana* 211-8k also appears to have a high level of phytol (**Figure 1-11**B) which is an important part of Vitamin K and Vitamin E, is used as a fragrance and has medicinal qualities such as antioxidant, anti-inflammatory, some antimicrobial activity and has been investigated as a drug for Schistosomiasis (de Moraes et al., 2014). Of the strains, *C. sorokiniana* 211-8k (UTEX 1230) has comparatively high amounts of these alternate compounds as percentage per cell which reduces the total amount of the other normal fatty acids (**Figure 1-11**).

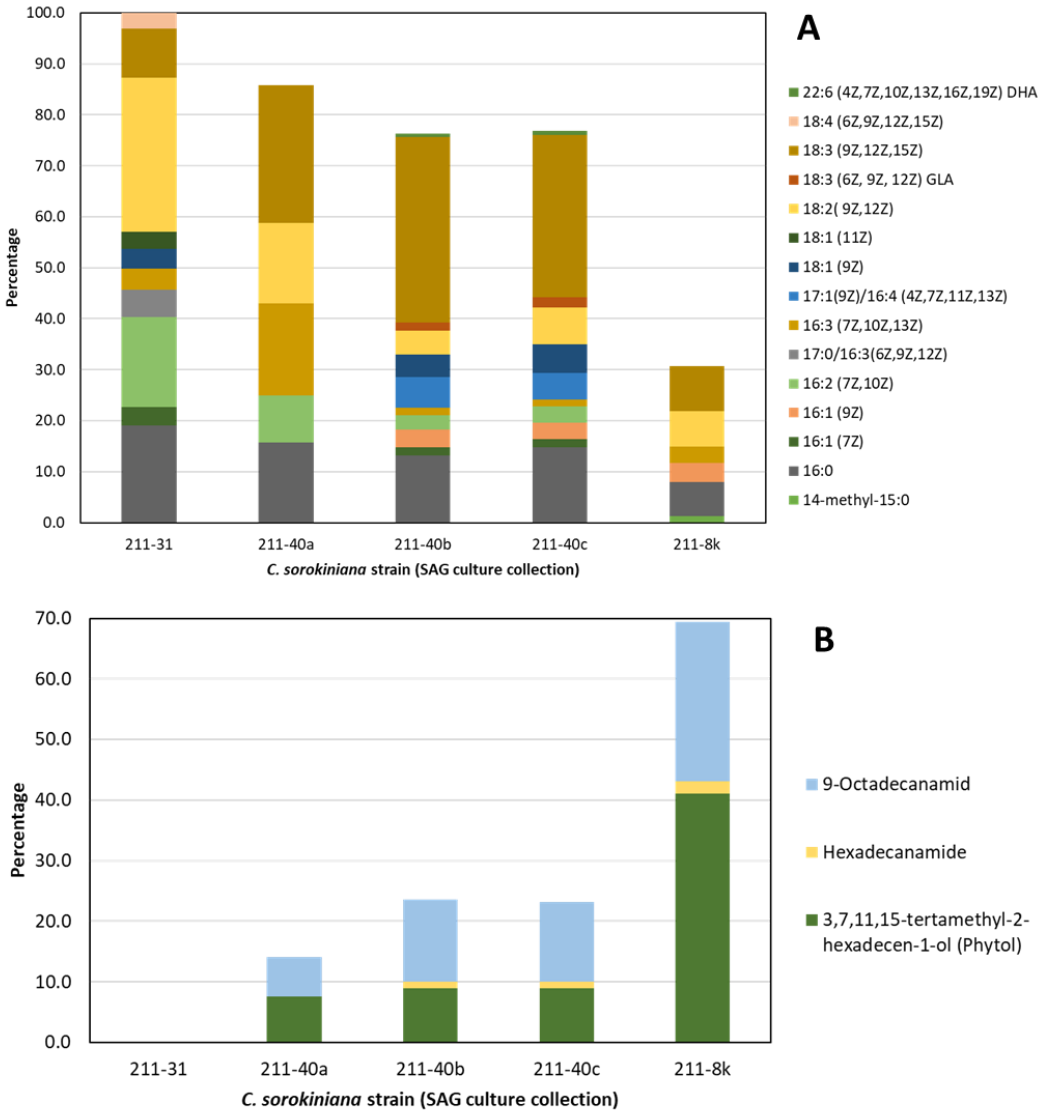


Figure 1-11 Fatty-acid profile of five *Chlorella sorokiniana* strains from the SAG culture collection. Data from supplementary data of Lang *et al*., 2011 (Lang et al., 2011) is used to make these graphs. For clarity the data is separated so **A.** Shows normal fatty-acids and **B.** shows alternative products characterised including primary fatty acid amides and Phytol. Strain 211-8k is the same as *C. sorokiniana* UTEX1230.

1.5.2 *Chlorella sorokiniana* UTEX 1230 characteristics and studies

The strain of *C. sorokiniana* used in this work is UTEX 1230 from the University of Texas Culture Collection (https://utex.org). It was first described in 1953 as an undesignated *Chlorella* strain Tx 71105 that had an optimal growth temperature of 39 °C (stable range 25.5 to 41.2) and reportedly the highest photosynthetic rate per

cell material of any organism at the time (Sorokin and Myers, 1953). Subsequent characterisation experiments were conducted under the designation "*Chlorella pyrenoidosa*, high-temperature strain 7-11-05" compared to the "Emerson" strain of *C. pyrenoidosa* (now *Chlorella emmersoni*) which could not withstand such growth conditions (Sorokin, 1959; Sorokin and Krauss, 1959) before assuming its correct taxonomic name of *Chlorella sorokiniana* Shihira and Krauss (Shihira and Krauss, 1965). Versions of *C. sorokiniana* UTEX 1230 are deposited in multiple culture collections under different names, for example as CCAP 211/8K (https://www.ccap.ac.uk) and SAG 211-8k (http://sagdb.uni-goettingen.de) (**Figure 1-12**) and therefore some publications may refer to these names.

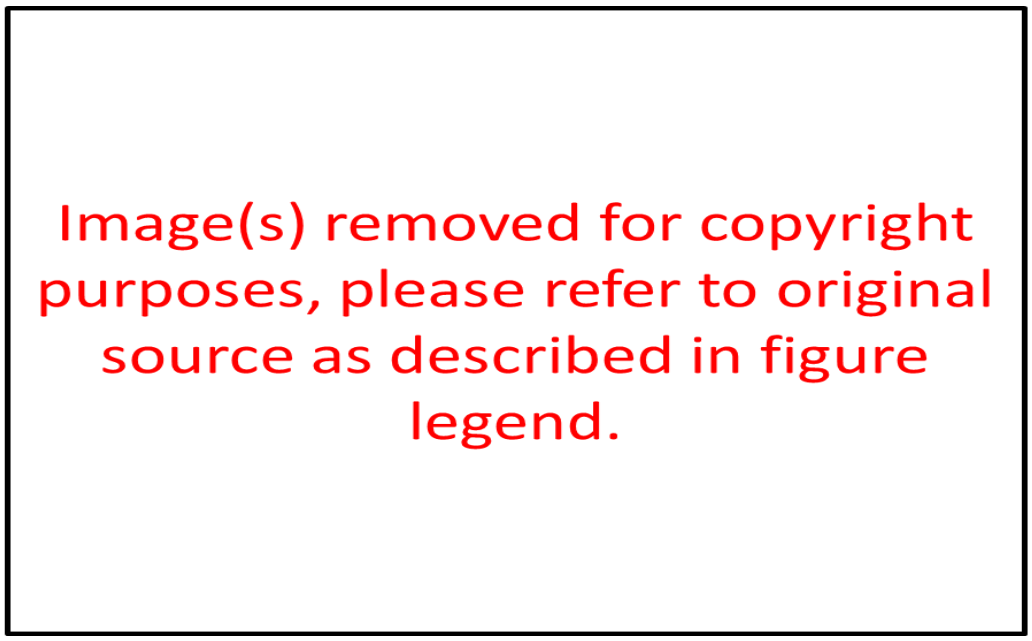


Figure 1-12 Micrograph of *Chlorella sorokiniana*. SAG 211-8k is the equivalent of CCAP 211-8k and UTEX 1230. This image is reproduced from the SAG culture collection (http://sagdb.uni-goettingen.de).

C. sorokiniana UTEX 1230 has been characterised as fast growing with doubling times reported of 2.6 – 5 hours (Li et al., 2014; Sorokin, 1959; Sorokin and Krauss, 1959). It is able to withstand temperatures up to 42°C (Li et al., 2013; Sorokin, 1959) and high light intensity (Cuaresma et al., 2009; Sorokin, 1959). It has a cell wall classified by the presence of glucosamine in the rigid wall and rhamnose as the hemicellulose sugar (Takeda, 1988). The wall contains no cellulose and is able to be

degraded by acetic acid (though there is some discrepancy with this observation) (Baudelet et al., 2017). Throughout the growth phase of *Chlorella* sp., the cell wall begins as single layered but then develops as a three-layered structure (Baudelet et al., 2017). This rigid cell wall and small size can be a problem for extraction of bio-products as it can be difficult to rupture (Lee et al., 2012).

Other physical characteristics include cells of varying (but relatively small) size of elliptical 3 x 2 to 4.5 x 3.5 µm cells increasing on size to 4.5-5.5 µm diameter on glucose, which also causes bleaching of the shallow bowl-shaped chloroplast (Shihira and Krauss, 1965) (**Figure 1-12**). A pyrenoid surrounded by a thin starch sheath is present within the chloroplast and is bisected by a double thylakoid membrane as seen in SEM images of strain 211-8k (SAG collection strain equivalent to UTEX 1230) (Ikeda and Takeda, 1995). This strain of *C. sorokiniana* is not sensitive to chlorovirus infection as tested under the name of 211-8k (Linz et al., 1999).

As for lipids, strain UTEX 1230 has been shown to accumulate high levels of TAGs under certain growth conditions such as heterotrophy on glucose (Rosenberg et al., 2014). These characteristics make it a highly attractive target for biotechnology and the production of biofuels or high value products. It has also received much interest as an industrial alga for bio-remediation of waste organics (Kobayashi et al., 2013) and flue gas (Kumar et al., 2014). The fatty-acid profile of *C. sorokiniana* will be discussed in detail in Chapter 3.

1.6 Progress in algal transgenics

To genetically engineer an organism there needs to be at least basic molecular tools available such as stable DNA transformation methods, good selection markers and suitable genetic parts such as promoters (Hlavova et al., 2015). Annotated genome sequences and a range of suitable regulatory sequences are also needed for more complex applications involving multigene expression and as a supply of endogenous regulatory sequences. There are currently only a few suitable microalgal species with these basic tools available. These include the model green alga *C. reinhardtii* which has the most extensive toolbox, the primitive red or green algae

Cyanidioschyzon merolae and *Ostreococcus tauri*, the diatoms *P. tricornutum* and *Thalassiosira pseudonana*, and the Eustigmatophyte *Nannochloropsis* sp. (Hlavova et al., 2015). The organism used in this project, *C. sorokiniana*, is not among these listed, though advancements have been made in other members of the *Chlorella* sp. genus in terms of transformation protocols, transcriptomics and genome projects as discussed below.

1.6.1 Microalgae transformation

Genetic manipulation in microalgae typically involves transformation of the nuclear genome, with chloroplast transformation possible for *C. reinhardtii* and a few other algal species, and the mitochondrial transformation reported only for *C. reinhardtii* (Spicer and Purton, 2016). Despite there being only a few microalgae with basic genetic toolkits, in fact over 40 species have reportedly been transformed successfully (Gangl et al., 2015). The DNA delivery techniques used successfully vary between species but can include microparticle bombardment, electroporation, glass-bead mediated, use of silicon-carbide whiskers, aminoclay nanoparticles, polyethelene glycol (PEG) mediate DNA uptake, episome transfer by bacterial conjugation and nuclear DNA insertion by *Agrobacterium*-mediated transformation (Charoonnart et al., 2018; Gangl et al., 2015; Hallmann, 2007; Karas et al., 2015; Liu and Chen, 2016). Microparticle bombardment is the most efficient and reproducible method but requires specialist equipment and expensive gold microparticles (Banerjee et al., 2016). The target genome of choice depends on the application and purpose of the transgenic strain.

1.6.1.1 Nuclear transformation

Transgene expression in the nucleus has advantages such as facilitating full eukaryotic post-translational modification of proteins including glycosylation, and stricter regulation of transgene expression by using inducible promoters which respond to a wide range of specific chemical or physiological factors (Rasala and Mayfield, 2015). Expressed proteins can also be targeted to specific parts of the microalgal cell: in *C. reinhardtii*, signal sequences have been successfully employed to target proteins to the cytoplasm, nucleus, chloroplast, or Golgi apparatus for

secretion (Lauersen et al., 2015). Secretion is especially useful for strategies where 'milking' of the algae for recombinant proteins in continuous cultivation is preferable and/or more cost effective (Akbari et al., 2014). Nuclear genomic engineering has advanced over the past decade in many organisms to include targeted specific sequence modifications, gene knock-outs or gene knock-ins by nucleases which have sequence specific DNA-binding modules. These include zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) or associated programmable guide RNAs (CRISPR/Cas) (Gaj et al., 2013). In microalgae, this technology is still yet to be fully established, with only a couple of model species reported so far for ZFN, TALENs and CRISPR (Ferenczi et al., 2017; Gimpel et al., 2015; Hlavova et al., 2015; Nymark et al., 2016). However, as work in this area progresses, there is much potential for it to facilitate precise metabolic engineering in microalgae with many industrial applications for such engineered strains.

1.6.1.2 Chloroplast transformation

For recombinant protein production, where microalgae are used as cell-factories, the chloroplast has some advantages over nuclear gene expression. These include the lack of gene silencing, the potential for poly-cistronic processing, the ability to target genes to specific places in the genome due to homologous recombination of exogenous DNA, higher levels of gene expression due to the ploidy of the chloroplast genome, more versatility to express genes from different organisms and lack of glycosylation (though in some cases glycosylation is a disadvantage depending on product purpose). The plastid can also serve as a compartment within the cell which can protect the recombinant products from degradation by host cytoplasmic proteases (Akbari et al., 2014; Charoonnart et al., 2018; León-Bañares et al., 2004; Specht et al., 2010). Disulphide bond formation and phosphorylation are also possible post-translational modifications, and there is the potential for light-inducible expression (Rasala and Mayfield, 2015).

1.6.2 Sources of genetic parts

Genetic parts include regulatory sequences, such as promoters, terminators and untranslated regions, and also coding sequences such as selectable markers and

reporters as well as those for producing a specific metabolite. Genetic parts may come from a variety of sources. Typically, algal regulatory sequences may be endogenous or from closely related species as they have more predictable effects than when using foreign DNA. However, the successful use of heterologous parts such as viral promoters, as was used in plants, can give rise to high levels of expression (Kay et al., 1987). For coding sequences, the source of the sequence information is more likely to be heterologous as the expression product will typically be a non-native protein such as a herbicide resistance gene or enzymes from foreign metabolic pathways.

For manipulation of the metabolism of the cell, upregulation of endogenous genes can be achieved by expressing additional copies of native genes. In the case of heterologous coding sequences, the use of codon optimisation of the sequence based on the codon bias of the desired host organism can increase expression (Heitzer et al., 2007). However, there is now increasing awareness that using the most frequent codons for all the amino acids within a coding sequence may not give the best expression profile because the speed at which the nascent amino acid chain emerges from the ribosome can affect its folding (Zhao et al., 2017). In fact the best strategy may be to match the codon usage to that of the native organism from which the protein originates, so if an organism uses a particularly rare codon at a certain position, it is best to mimic this in the heterologous host as well, especially if the rarity is conserved across multiple species (Quax et al., 2015). For regulatory sequences, there is also work on creating entirely synthetic parts based on rational minimal design based on analysis of motifs in the host organism (Scranton et al., 2016). Advantages include that the synthetically designed parts are often smaller as they can contain condensed versions of endogenous parts (Liu and Stewart, 2016; Vogl et al., 2014). Smaller constructs are generally easier to manipulate in cloning cycles and transformation into the host. They also avoid issues with cross-talk from host cell control.

1.6.2.1 Available algal genomes as sources of endogenous parts

Endogenous regulatory sequences are commonly used for work in the model green alga *C. reinhardtii* (Gangl et al., 2015). The use of endogenous sequences generally requires a genome sequence of the organism. Alternately, a region of interest can be isolated from a genomic extract by PCR with degenerate primers designed using the genome of a closely related organism, but this is unreliable depending on how conserved is the sequence. However, there are now a multitude of algal genomes available from diverse phylogenetic groups including many green algae and red algae (Banerjee et al., 2016; Hallmann, 2007), which gives more possibility for the use of endogenous parts.

At the beginning of this project there was not a sequenced genome for *C. sorokiniana*. However, now there is much work in this area. Multiple nuclear genome sequences are available for *Chlorella* species in online databases. Annotated genomes have been published for *Chlorella protothecoides* sp. 0710, *Chlorella variabilis* NC64A and *Chlorella pyrenoidosa* FACHB-9 and *C. sorokiniana* (**Table 1-3**) (Blanc et al., 2010; Fan et al., 2015; Gao et al., 2014). Of these, *C. pyrenoidosa* FACHB-9 is the most closely related to *C. sorokiniana*. The variation in the size of these genomes and the number of predicted genes suggests large amount of diversification (Fan et al., 2015) but they can still be a useful reference. However, care must still be taken when drawing assumptions for *C. sorokiniana* UTEX 1230 since lipid metabolism and related pathways of interest for this project are the least constrained (Fan et al., 2015). The *C. sorokiniana* UTEX 1230 genome came out of a large body of work by the National Alliance for Advanced Biofuels and Bioproducts (NAABB) consortium funded by the US Department For Energy (NAABB, 2014). The draft sequences are available online from the Los Alamos National Laboratory website https://greenhouse.lanl.gov/organisms/ (LANL, 2014) (**Table 1-3**). Some genomes are accompanied with detailed transcriptomic data (**Table 1-3**) which provides important information for informing metabolic engineering strategies (Fan et al., 2015; Gao et al., 2014). In some cases, transcriptomic and proteomic analysis were carried out before a genome was available, as is the case for *C. vulgaris* UTEX 395 (Guarnieri et

al., 2013, 2011). There are also some cases where just the chloroplast genome is available e.g. *C. vulgaris* C-27 (Genbank accession AB001684).

Species	Genome characteristics and notes	Reference/location
***Chlorella protothecoides* sp. 0710**, a promising candidate for biofuel production	22.9 Mbp, GC 63%, annotated, transcriptomic data available	(Gao et al., 2014; Wu et al., 2015)
***Chlorella variabilis* NC64A**, a model in the study of eukaryotic algae chlorovirus virus-host dynamics	46.2 Mbp, GC 67%, annotated, transcriptomic data available, genome scale metabolic model available	(Blanc et al., 2010; Juneja et al., 2016; Rowe et al., 2014)
***Chlorella pyrenoidosa* FACHB-9**, an important strain of interest for protein, starch and lipids	56.8 Mbp, GC 66.6%, annotated, transcriptomic data available	(Fan et al., 2015)
***Chlorella sorokiniana* UTEX 1230**, a fast-growing productive strain for biofuel production	58.5 Mbp, annotated, no paper	(LANL, 2014; NCABB, 2016)
***Chlorella* sp. 1228**, a clonal isolate 'Phycal' derived from *Chlorella sorokiniana* UTEX1230	61.4 Mbp, annotated, no paper	(LANL, 2014; NAABB, 2014)
***Chlorella* sp. DOE1412**, a unique lab isolate with a closest genetic relationship to *Chlorella sorokiniana*	59.3 Mbp, annotated, no paper	(LANL, 2014; NAABB, 2014),
***Chlorella vulgaris* UTEX 395**	63 Mbp, no paper, metabolic model constructed, transcriptomic and proteomic data available	(Guarnieri et al., 2013, 2011; Huelsman, 2015); NCBI BioProject PRJNA278897

Table 1-3 Nuclear genome sequencing efforts for selected *Chlorella* species. The NCBI accession numbers are given. This list is not exhaustive. Those with the reference LANL 2014 are publicly available at: https://greenhouse.lanl.gov/greenhouse/organisms/

1.6.2.2 Reporters and selectable markers

Microalgal selectable markers typically confer resistance to an antibiotic or herbicide such as zeocin, G418, chloramphenicol, hygromycin, Basta or Norflurazon. Reporters typically encode a product such as a fluorescent protein that can be detected at a graduated level. This can allow for determination of optimal integration sites and assaying the performance of regulatory sequences such as promoters. The reporter gene encoding the *E. coli* enzyme β-glucoronidase (GUS) is commonly used in plants and has been used in microalgal work as well (Kumar et al., 2004; Li et al.,

2009; Wang et al., 2007). Green fluorescent protein (GFP) or enhanced GFP have been successfully expressed and detected in multiple microalgae including the models *C. reinhardtii* and *P. tricornutum* (Kumar et al., 2004). There are now comprehensive lists of reporters and selectable markers, and records of their use in different algae (Bashir et al., 2016; Gangl et al., 2015; Hallmann, 2007).

There is also interest in developing marker-free or positive selection systems, also known as self-cloning or cis-genesis, because these are more amenable to use industrially. Genetically modified algal strains can only generally be cultivated in enclosed photobioreactors rather than outside, depending on the regulations defining a country (Leu and Boussiba, 2014). This can lead to reduced economic viability of these strains because photobioreactors have additional considerations for capital and operational expenditure than open pond systems (Leu and Boussiba, 2014). These techniques started with research in plant systems which face the similar regulatory setbacks (Puchta, 2003). In one example, a system for removal of a marker gene after selection by subjecting plants to heat shock was developed in grapevine (Dalla Costa et al., 2015).

For microalgae, marker-free nuclear transformation has been achieved in *C. vulgaris* using linear gene expression cassettes containing DNA-binding and modifying domains of the Simian virus 40 (SV40) large T antigen (Gomma et al., 2015). For transformation of the chloroplast, marker-free systems typically involve the use of mutant 'recipient' strains which are deficient in some way and then selection is a positive process based on restoration of function. For example uracil auxotrophy was a successful strategy used in the green alga *Pseudochoricystis ellipsoidea* (Kasai et al., 2015). In *C. reinhardtii,* the mutant 'Fud7' recipient strain can be used, which cannot grow autophototrophically due to deletion of the psbA gene encoding the core photosystem II subunit, D1 (Bertalan et al., 2015). These complementation approaches work well but can be inaccessible for non-model species because the generation and isolation of mutant lines containing the appropriate phenotype for restoration is a lengthy and costly process (Walker et al., 2005). Additionally, the dependence on homologous recombination to restore the function in the original gene means these approaches are generally unsuitable for nuclear transformations

because homologous recombination is much rarer in the nucleus of most microalgae (Doron et al., 2016).

There are also safety concerns with the possibility of unwanted horizontal gene transfer or release of foreign DNA into the environment from genetically modified organisms. A valuable strategy against this was developed by making the transferred genes unreadable to most microorganisms by including stop codons in the DNA which would prohibit translation of the full peptide chain. This was possible by taking advantage of an unused 'stop' codon in the *C. reinhardtii* chloroplast and reassigning the codon to a specific amino acid with a modified tRNA carrier (Young and Purton, 2015).

Although these marker-free approaches are valuable, and no doubt improve the cost of selection, it could be considered ethically questionable as to whether it is appropriate to try and circumvent regulations by defining different terms such as cis-genesis or intra-genesis, where strains are engineered only with genetic elements from phylogenetically related species or a sexually compatible donor rather than foreign DNA (transgenesis). The bottom line is that all these technologies require the transformation of the organism, no matter what the source DNA is or whether just a single base-pair has been edited. For example, there are many undesired integrations and modifications that can happen during transformation and DNA editing technologies. For example, aberrant integration of non-specific vector-backbone DNA during *Agrobacterium*-mediated transformation has been documented to occur in plants between 38-85% of cases (Gelvin, 2003; Rommens, 2004; Windels et al., 2008) and chromosomal *A. tumefaciens* DNA may be present in one in every 250 transgenic plants (Ulker et al., 2008). For gene editing, the off-target effects of the CRISPR/Cas9 system are well-known (Zhang et al., 2015). Therefore, unless the genome of a genetically engineered organism is entirely sequenced, there is uncertainty as to whether the organism is truly free of un-intended DNA parts such as that of cloning vector backbones (which may contain antibiotic resistance markers) or any off-site rearrangements or modifications. Arguably, traditional mutagenic and breeding techniques such as UV irradiation also cause unknown aberrations in DNA. However, these concerns should not be ignored, instead they should be embraced

positively as part of an increasing awareness of the importance of the governance of science through responsible innovation, on which there is much recent conversation (Li et al., 2015; Macnaghten, 2016).

1.7 Thesis aims and summary

The work presented in this thesis aimed to explore the proof-of-concept production of LC-PUFAs in the industrially relevant microalgae *Chlorella sorokiniana* UTEX1230, which requires a genetic engineering approach to add the relevant genes to the pathway. The work is therefore divided into three parts:

- The first results chapter (chapter 3) investigates factors which affect *C. sorokiniana* growth and lipid profile by assessing different nutrient regimes and their subsequent impact on fatty-acid profile, growth and neutral lipid accumulation. One focus of this chapter is to determine an appropriate growth condition for maximising C18:3n3 (ALA) which is the precursor in the omega-3 LC-PUFA pathway.
- The second results chapter (chapter 4) aims to investigate transformation strategies for *C. sorokiniana* with a focus on optimisations of an *Agrobacterium*-mediated method
- The third results chapter (chapter 5) aims to address the need for genetic parts in algal biotechnology by identifying novel regulatory sequences from chlorovirus genes
- The fourth results chapter (chapter 6) incorporates these sequences and others into a parts library compatible with standardised Golden Gate cloning and investigates the ability of the chlorovirus regulatory sequences to drive gene expression in *C. sorokiniana* and *C. reinhardtii*.

CHAPTER 2 MATERIALS AND METHODS

2.1 Bacterial and Algal strains used in this work

2.1.1 *Chlorella sorokiniana*

Chlorella sorokiniana UTEX 1230 had been previously obtained from the culture collection at the University of Texas (UTEX: www.utex.org). Culturing conditions are described later.

2.1.2 *Chlamydomonas reinhardtii*

Chlamydomonas reinhardtii cw15 cc-5155 mt+ had been previously obtained from the *Chlamydomonas* Resource Centre (https://www.chlamycollection.org/).

2.1.3 *Agrobacterium tumefaciens*

Agrobacterium tumefaciens strain used was EHA105 (Chromosomal Background C58, Ti plasmid pEHA105 (pTiBo542ΔT-DNA)) (Hood et al. 1993).

2.1.3.1 Culturing conditions

A. tumefaciens was cultured in liquid or solid (10 g/L agar) YEN-B medium (7.5 g/L Difco Bacto™ yeast extract, 8 g/L Fluka Nutrient broth, pH 7.5 with NaOH or HCl) at 28 °C. Culture vessels were sterile and liquid cultures were shaken at ~150 rpm and contained at least 4/5ths volume as headspace for aeration. Temperature was checked regularly to ensure it did not rise above 28 °C, since this can cause loss of the Ti plasmid. Agar plates and liquid cultures were supplemented with antibiotics at appropriate with concentrations of 50 µg/mL for kanamycin.

For long term storage, 500 µl of an overnight culture (with appropriate antibiotics) was supplemented with glycerol to a final concentration of 25% in screw-top 1.5 mL tubes and stored at -80 °C.

2.1.3.2 Preparation of competent cells

A stock of the required wild-type *A. tumefaciens* strain was streaked on YEN-B agar supplemented with appropriate antibiotics and incubated for 28 °C for 36 – 48 hours or until single colonies appear. A single colony was picked into 10 mL YEN-B broth supplemented with appropriate antibiotics and shaken ~19 hours at 28 °C before seeding 8 mL into 200 mL YEN-B plus appropriate antibiotics and incubating for ~20 hours. The culture was incubated on ice for 10-20 minutes before splitting into 4 x 50 ml tubes and harvesting by centrifugation (3500 g) in a chilled (4 °C) centrifuge and decanting supernatant. Each cell pellet was resuspended in 25 ml cold 100 mM $CaCl_2$ by pipetting. The harvesting was repeated and cells resuspended in 1 ml cold 20 mM $CaCl_2$ before aliquoting 100 µl into sterile 1.5mL tubes and storing at -80 °C.

2.1.3.3 Transformation using freeze-thaw protocol

Approximately 200 ng of the desired plasmid DNA was mixed with a vial of chemically competent *Agrobacterium tumefaciens*, thawed by hand and flicked for 5-10 seconds by hand to ensure mixing. The mixture was then incubated in liquid nitrogen for 5-10 minutes before incubating at room temperature for 5-10 minutes. The cells were transferred to 1 ml YENB broth in 30 ml universal tube and recovered for 4-16 hours in a shaker (~100 rpm) at 28 °C before plating onto YENB agar plates supplemented with appropriate antibiotics and incubating plates at 28 °C until single colonies appear (~48 hours). Confirmation of the presence of plasmid in single colonies was performed using colony PCR (polymerase chain reaction) and stocks were made of colonie(s) containing the confirmed construct.

2.1.4 *Escherichia coli* cloning host

DNA manipulation and cloning was carried out in *Escherichia coli* strain DH5α (genotype: F-, *(φ80lacZΔM15), Δ (lacZYA-argF)*U169, *deoR, recA1, endA1, hsdR17(r_k-,m_k+), supE44, thi-1, gyrA96, relA, λ*).

2.1.4.1 Culture growth and maintenance

Liquid cultures were grown in autoclaved LB medium (10 g/L Difco Bacto™ tryptone, 5 g/L Difco yeast extract, 2 g/L NaCl) at 37 °C with shaking at 150-200 rpm. Where solid cultivation was required, LB-medium was supplemented with agar to 1.5%. Media was supplemented with antibiotics at appropriate: kanamycin at 50 µg/ml, ampicillin at 100 µg/ml. For long term storage strains, 500 µl of an overnight culture was supplemented with glycerol to a final concentration of 25% and stored at -80 °C.

2.1.4.2 Preparation of plasmid DNA

Plasmid DNA was purified from *E. coli* using a kit (Thermo GeneJet). 5 mL cultures supplemented with appropriate antibiotic were grown in 50 mL or 30 mL sterile plastic tubes overnight and harvested by centrifugation at 4000 *g*. Supernatant was discarded and manufacturer protocol was followed.

2.1.4.3 Preparation of *E. coli* chemically competent cells

DH5α *E. coli* from a glycerol stock was streaked out on an LB plate containing no antibiotics and left overnight at 37 °C. A single colony from this plate was grown overnight, shaking at 37 °C in liquid LB containing no antibiotics. From this culture, 0.5 mL was added to 50 mL liquid LB containing no antibiotics in a 250 mL conical flask or 0.125 mL culture to 12.5 mL LB in 4 x 50 mL centrifuge tubes which were grown at 37 °C, shaking, for 2.5 hr. The culture(s) were then cooled on ice for 15 minutes before pelleting the cells in 2 x pre-cooled 50 ml centrifuge tubes at 4 °C for 5 minutes at 3736 g. The supernatant was discarded and 10 ml cold 50 mM $CaCl_2$ was added to the tubes on ice and gently resuspended using a stripette pipette. The cell suspensions were left on ice for 30 minutes before spinning as previously described and removing the supernatant and adding 1.5 mL 50 mM $CaCl_2$ and resuspending as previously described. After pooling cells into one tube and adding 1.4 mL of sterile 50 % glycerol, 220 µl of cell suspension was added to pre-cooled 1.5 ml cryovials before freezing and storing at -80 °C.

2.1.4.4 Heat-shock transformation of *E. coli*

Plasmid DNA was added to 50 or 100 µl chemically competent *E. coli* cells to no more than 10 % of the total volume of cells. The mixture was incubated on ice for 10-30 minutes. Cells were given a heat shock by placing in a 42 °C water bath before returning to the ice and adding 500 or 1000 µl LB (for cell starting volumes of 50 or 100 µl respectively). Cells were then recovered by incubating at 37 °C for 20-60 minutes before plating out on LB agar containing the appropriate antibiotics. A maximum volume of 200 µl was plated. If deemed necessary, the recovery culture was concentrated by spinning at 13,000 g in a microfuge and resuspending in the required amount of LB.

2.2 Molecular Biology (DNA analysis and manipulation)

2.2.1 DNA Sequence manipulation software

Primer design and manipulation of DNA sequences in silico was done using the Benchling online software which can be found at https://benchling.com.

2.2.2 PCR and Colony PCR

Polymerase Chain Reaction (PCR) was used to confirm plasmid presence, genomic integration and to amplify desired DNA fragments for Golden Gate cloning. A typical PCR reaction used Phusion or Q5 DNA polymerase from NEB (New England BioLabs) and reactions vessels were set up according to manufacturer recommendations. Primer details can be seen in the appendix section A.I pg 316, and annealing temperatures used were those recommended by the manufacturer depending on the polymerase used.

For direct confirmation of plasmid within bacterial cells, colony PCR was carried out using cells as a source of template DNA. A colony was picked into a 200 µl PCR tube before adding 20 µl Taq Polymerase master mix made according to NEB guidelines with the desired primers (see appendix). The PCR programme was run according to NEB guidelines including the additional denaturation step of 98 °C for 6 min to aid cell lysis. If necessary, a copy plate was done of the colony before touching

the colony to the PCR tube. For *Agrobacterium*, 1-2 µl of an overnight culture was occasionally used as the template instead of a colony.

2.2.3 Restriction Enzyme Digestion

Restriction enzymes were ordered from NEB and reactions were assembled according to manufacturer guidelines. In general, reactions were 50 µl in total volume, assembled on ice in sterile 1.5 mL tubes and contained CutSmart™ buffer, autoclaved dH2O, template DNA at desired concentration and 1 or 2 restriction enzymes.

2.2.4 Ligation

Ligation was carried out using T4 DNA ligase (NEB) according to manufacturer guidelines: 20 µl reactions with required amount of vector and insert and incubation for 10 mins at room temperature. For some reactions the DNA was purified using agarose gel electrophoresis and a Gel Extractions Kit (Thermo) or a PCR purification Kit (Thermo) as required.

2.2.5 Golden Gate DNA Assembly Reactions

2.2.5.1 Level 0 Parts library

Level 0 parts were made by PCR from various templates (see section 6.1.2). Q5 of Phusion DNA polymerase was used (NEB) using manufacturer recommendations. Annealing temperatures and primers can be found in the appendix. The primers contained overhangs consisting of the converging *BsaI* restriction enzyme sites followed by the appropriate 4 bp DNA overhang for the appropriate part type (Patron et al., 2015). The PCR products were cloned into pJET 1.2 vector as the Level 0 plasmid which has Ampicillin as the resistance marker.

2.2.5.2 Assembly of Level 0 parts into expression vectors

To assemble the Level 0 parts into the modified pCAMBIA-2300 based Level 1 acceptor vector pCor-J23-div (described in section 6.1.4), reactions were set up as follows: Reactions contained: 2 µl CutSmart buffer (NEB), 2 µl ATP (10 mM), 1 µl T4

ligase (NEB), 0.5 µl *BsaI*-HF (NEB), Level 0 part plasmids, destination vector DNA and dH2O to 20 µl. Enzymes were added last. The amount of DNA was calculated according to the length of the part: the destination vector concentration was set to 100 ng and the insert (Level 0)/acceptor (destination) plasmid ratio was set to 2:1. The ng of insert plasmid needed was then calculated according to the following formula: ng level 0 plasmid = desired insert/vector ratio * mass of vector (g) * ratio of insert to vector lengths and is from NEB website. Assembled reactions were then placed in a thermocycler (with heated lid at 105 °C) on the following programme: 10 min 37 °C, 10 min 16 °C 3 times, 10 min 37 °C, 20 min 80 °C, hold 16 °C. The reaction was stored at -20 °C or used immediately for transformation.

For transformation, 2-5 µl of the reaction was then used to transform chemically competent *E. coli* DH5-alpha cells using a standard heat-shock transformation protocol on selective kanamycin (50 µg/ml) plates, incubated overnight at 37 °C. Plates were examined and colourless colonies were picked for colony PCR analysis to confirm the presence of the assembled expression cassette in the vector. If necessary plates were left at room temperature for ~6 hours to allow RFP colour to accumulate more to ensure they were not picked. Correctly identified clones were then grown up and plasmid DNA isolated using a miniprep kit from Thermo, eluting in 30 µl water and storing at -20 °C.

Modifications to improve the efficiency of the Golden Gate reaction for these parts are described in the results section 6.1.4.

2.2.6 Agarose Gel Electrophoresis

A 50 mL agarose gel at 1-2% in TAE buffer was prepared using a microwave and allowed to cool before addition of Ethidium Bromide to a final concentration of 15 µg/ml. Gels were run at 80-100v for 35-60 min and visualised under UV light.

2.2.7 Sequencing

DNA sequences were sent for sequencing at an external company Source Biosciences (https://www.sourcebioscience.com).

2.2.8 Southern Blot

The southern blot was carried out in collaboration with Noreen Hiegle.

2.2.8.1 Blot preparation

The southern blot membrane was prepared previously by previous lab member Noreen Hiegle using standard molecular biology techniques. In summary, a genomic prep of *C. sorokiniana* was carried out and digested with NcoI or XhoI restriction enzymes and ran on an 0.8 % (w/v) agarose gel with Ethidium Bromide at 15 µg/ml, then transferred to neutral nylon Hybond N (Amersham Biosciences) membrane using standard southern blot procedures. Before blotting the gel had been imaged with a ruler as a reference for the ladder. The blotted membrane was stored dry at room temperature between 2 sheets of Whatman 3mm filter paper until hybridisation stage.

2.2.8.2 DNA probe

The sequence design of the DNA probe was carried out by previous lab member Noreen Hiegle. The designed probe was prepared by PCR (F primer 5'-AGGAGCAGGACTAACCGA-3', R primer 5'-ATCTGATCCAAGCTCAAGC-3') from the RbleC23 vector. 5 ng of RbleC23 template was added to 2.5 µl forward and reverse primer (10 mM stock), 10 µl HF or GC Buffer (NEB), 1 µl dNTP mix (10 mM stock), 0.5 µl Phusion HF (NEB) and dH_20 to 50µl in a 200 µl PCR tube. An additional no primer reaction was used as the negative control. Reaction was run in a thermo cycler with the following programme: 1) 98 °C 30 s, 2) 98 °C 10 s, 3) 60 °C 30 s, 4) 72 °C 15 s, 5) Repeat step 2-4 35 times, 6) 72 °C 600 s. A gel extraction was performed using a kit (Thermo) according to the manufacturer instructions and final DNA concentration measured using a NanoDrop ND 2000c spectrophotometer. The gel-extracted DNA fragment was DIG-labelled according to manufacturer instructions in the DIG High Primer DNA Labelling and Detection Starter Kit II (Roche) including an additional purification step using a PCR purification kit (Thermo).

2.2.8.3 Hybridisation

Hybridisation protocol was carried out according to manufacturer protocol. Using the formula provided for the full-length probe (532 bp), hybridisation temperature was calculated to be 48 °C. Hybridisation was performed overnight for ~ 14 hrs in a heat-sealed plastic bag in a rotating oven (HYBAID MINI 10) using the recommended 25 ng/ml concentration of DIG-labelled DNA probe. High-stringency wash was 0.1x SSC + 0.1% SDS at 68 °C as recommended for high homology and GC% probes. Chemiluminescent detection was carried out as per manufacturer instructions and subsequently the membrane was exposed to X-RAY film (GE Healthcare) for 20-40 min.

2.3 Growth and Handling of microalgal strains

2.3.1 Growth media recipes

2.3.1.1 TAP medium

The original Tris-Acetate-Phosphate (TAP) medium is from the 1960s (Gorman and Levine, 1965; Sueoka, 1960). Acetate is provided as a carbon source at 17.4 mM, and nitrogen as ammonia at 7.48 mM. Each litre of TAP medium contains 2.42 g/L Trizma Base, 1 mL 1 M (K)PO_4 pH 7.0 (1 M Monobasic Potassium Phosphate K_2HPO_4 titrated to pH 7.0 with Dibasic Potassium Phosphate KH_2PO_4), Glacial Acetic acid to pH 7.0 (=17.4 mM final concentration), 25 mL of 4× Beijerinck salts (0.3 M NH_4Cl (=7.48 mM final 1x concentration), 14 mM $CaCl_2$·2H2O and 16 mM $MgSO_4$·$7H_2O$) and 7 mL combined trace element solution from Kropat *et al* (Kropat et al., 2011): 25 mM EDTA-Na_2, 28.5 µM $(NH_4)_6Mo_7O_{24}$, 0.1 mM Na_2SeO_3, 2.5 mM $ZnSO_4$ in 2.75 mM EDTA, 6 mM $MnCl_2$ in 6 mM EDTA, 20 mM $FeCl_3$ in 22 mM EDTA and 2 mM $CuCl_2$ in 2 mM EDTA. Stocks of trace element solution, 4x Beij. salts and 40x remaining components are autoclaved separatley and stored at 4 °C before being used to make up a 1× solution which is then autoclaved. When used as a solid medium, 1× TAP is supplemented with agar at 2% before autoclaving.

2.3.1.2 TP medium

Tris-Phosphate (TP) medium, sometimes referred to as TAP-minimal medium, is identical to TAP medium except the Glacial Acetic Acid component is omitted and the media is instead brought to pH 7.0 using HCl.

2.3.1.3 Carbon source variation

Where alternative carbon sources or concentrations to the default TAP medium were required, TP medium was used and was supplemented with the required amount of glucose, xylose, glycerol or Glacial Acetic acid. pH was adjusted as necessary with NaOH or HCl. Exact concentrations are described in the experimental chapters.

2.3.1.4 Nitrogen source variation

In some experiments, the nitrogen source was varied. In this case, the 4× Beijerinck solution is modified to contain the desired amount of ammonium as Ammonium Chloride (NH_4Cl), or nitrate as Sodium Nitrate ($NaNO_3$). Exact concentrations are described in the experimental chapters.

2.3.1.5 BBM

Bold's Basal Media is another commonly used freshwater green algae growth medium. BBM was ordered as Bold modified basal freshwater nutrient solution (modified BBM, Sigma #B5282), 20x sterile liquid stock. BBM contains nitrate as a nitrogen source and no carbon source. In this work, BBM is supplemented with glucose at 2 to 10 g/L as specified in the experimental chapters.

2.3.2 Maintenance of working stocks on Agar Plates

C. sorokiniana strains were maintained by re-streaking a loopful of cells on fresh TAP agar (2%) plates every 4-6 weeks. All propagations were carried out in a flow hood with either sterile plastic loops of 10 µl volume or metal loop sterilised by Bunsen burner flame.

2.3.3 Algal growth in liquid shake flasks

Standard growth conditions were mixotrophy in TAP medium at 25 °C, constant illumination via a white fluorescent light strip at ~150 µmol photons $m^{-2}s^{-1}$ and shaking at 120 rpm (Innova 44, New Brunswick Scientific). Starter (seed) cultures were 20-25 mL TAP in a 50 mL flask or 25-50 mL in a 100 mL flask. Starter cultures were used to inoculate larger cultures of 200 mL in 500 mL flask or 100 mL in 250 mL flask in media as specified in the experimental chapters. In general, experimental cultures were grown in triplicate at a minimum and flask positions were semi-randomised in the incubator shaker to minimise error from uneven light or temperature distribution within the incubator.

2.3.4 Algal growth in the Algem Photobioreactor

The Algem photobioreactor allows controlled growth of two flasks in parallel (**Figure 1-1**). 1 L flasks with 400 mL working volume were grown under 200 µmol photons $m^{-2}s^{-1}$ PAR (Photosynthetically Active Radiation: wavelength 400-700 nm) unless otherwise specified. Temperature was also controlled as specified. The Algem photobioreactor monitors OD at 740 nm automatically. Data was exported to and analysed in the software Microsoft Excel 2016.

Figure 2-1 The Algem Photobioreactor.

2.3.5 Sampling algal cultures

Samples were taken from flasks using sterile Stripettes of appropriate volume and transferring to a sterile container of required volume of either universal tube 30

mL, 7 mL, and conical tubes 15 mL or 50 mL. For smaller volumes a pipette was used and the neck of flask flamed if necessary.

2.3.6 Algal Growth measurements

Algal growth was measured as OD at 750 nm in a spectrophotometer (UNICAM UV vis). The culture was diluted until it gave a reading below one and the correct OD reading was gained by multiplying by the dilution factor. For manual cell counts a culture was diluted and applied to a haemocytometer and counted under a light microscope. Dilution was done to 100-200 cells per side of haemocytometer for accurate counting which corresponds to an OD_{750} of ~0.1.

2.3.7 Preparation of cells for long-term storage in Liquid Nitrogen

A cryogenic freezing container (Mr Frosty™ Nalgene™) was filled with isopropanol as per manufacturer instructions and pre-chilled. 5 mL TAP medium was inoculated with a large loop of cells from a freshly growing algal plate stock and incubated for 10 min shaking at 25 °C until no clumps appeared. 0.75 mL of the algal culture is added to cryovials containing 0.75 mL TAP and 10 % methanol (final concentration of MeOH is 5%) and placed in the cryogenic freezing container at -80 °C for 1.5 hr. Tubes are then transferred to Liquid nitrogen for long term storage.

2.3.8 Testing of putative *C. sorokiniana* transformants using spot tests

2.3.8.1 Antibiotic stocks

The antibiotics used for *C. sorokiniana* were G418 (Gibco Genetecin) and Zeocin (A phleomycin based antibiotic, Invivogen™ or Invitrogen™). Liquid working stocks stored at -20 °C according to manufacturer recommendations.

2.3.8.2 Spot tests

For testing of historical transformants: 3 mL TAP was inoculated with a small loop-full of cells from a TAP (Tris-Acetate-Phosphate) agar plate stock in a 7 mL sterilin for each mutant, then incubated with light and shaking at 120 rpm, 25 °C for ~20 hrs with the aim of reaching early-log or log phase. If any cultures grew to early-

stationary after 20 hrs according to OD_{750} measurements, cultures were diluted and then regrown for ~ 5 hrs until reaching an OD_{750} of ~1. All cultures were then diluted to an OD_{750} 0.85 before diluting 1/5, 1/25, 1/125, 1/100 for plating on antibiotic plates, and additional dilutions of 1/12500 and 1/125000 for TAP only plates. 10 µl was plated out in spots on TAP agar (2%) plates prepared with 100, 350, 400 or 450 µg/ml G418; 250, 300 or 350 µg/ml Zeocin; or no antibiotic as positive control. Plates were incubated with light at 25 °C for 8 days before imaging with a scanner.

Alternatively, for spotting for antibiotic sensitivity tests of wild-type when precise cell number was controlled, a 30 ml culture in a 100 ml flask was grown for ~30 hours to log phase before diluting to the cell number required and plating out 10 µl spots on the agar plates supplemented with appropriate antibiotics.

Spot tests were also prepared from cells directly from an agar plate stock or from single colonies on a plate. In this case a small loop of the cells was resuspended in 100 µl TAP medium. 10 µl of this neat suspension was spotted on agar plates containing the appropriate antibiotics or a no-antibiotic control. Spots were also done for 1:11 and 1:110 dilutions. ~66 µl of the neat suspension was used to prepare genomic DNA preparations as discussed in a different section.

2.3.9 Genomic DNA preparation of algal cells

One method used Chelex-100 resin (BioRad) for genomic DNA extraction (Wan et al., 2011). The cell suspension (between 5×10^5 and 1.5×10^7 cells as in (Wan et al., 2011) was spun at 10,000 g for 1 minute to pellet the cells. Supernatant was discarded and the pellet resuspended and vortex in 50 µl of 6% w/v Chelex-100 solution in dH_2O. The tube was heated for 100 °C for 10 minutes then cooled to 4 °C on ice. The sample was then vortexed and spun for 10,000 g for 1 minute before removing the supernatant and transferring it to a clean tube where 1-5 µl was used directly in a PCR reaction or stored at -20 °C until use.

An alternative method was based on boiling directly in the HF PCR buffer for Phusion DNA Polymerase (NEB) (Liu et al., 2014a). 1 or 2 µl of a cell suspension was

added to 10 µl of 2 x HF buffer in a PCR tube and mixed before boiling for different lengths of time (see results) before adding 11 µl of mastermix described in **Table 2-1**.

Item	**Volume for 1 master mix reaction (ul)**	**Concentration in final ~20 µl PCR reaction**
10 mM dNTP	0.4	200 µM
10 µM Forward primer	1	0.5 µM
10 µM R primer	1	0.5 µM
Phusion DNA polymerase	0.2	0.4 U
dH_2O	7.4	To 10 µl (the remaining volume is from the pre-boiled cells)
2x HF buffer	1	1x

Table 2-1 PCR master mix for *Chlorella* colony PCR direct in PCR buffer. From Liu *et al*. (Liu et al., 2014a).

2.4 Growth of *Chlorella sorokiniana* for lipid experiments

2.4.1 Comparing two different media

Replicate (n=4) 200 mL cultures in a 500 mL Erlenmeyer flask contained either TAP (tris-acetate-phosphate) media or Bold modified basal freshwater nutrient solution (modified BBM, Sigma #B5282) with or without supplementation with 10g/L glucose. Flasks were inoculated at a density of 0.6-1 million cells/ml from a pooled 20 mL starter cultures and incubated at 25 °C, 120 rpm (Innova 44, New Brunswick Scientific) and 150 µmol photons $m^{-2}s^{-1}$ white light from above (fluorescent strips). Heterotrophic flasks were covered in tin-foil to deprive the culture of light, mixotrophic culture had tinfoil over the cotton wool bung only. Starter cultures had been inoculated from an agar plate stock and incubated at 25 °C, 120 rpm, 150 µmol photons $m^{-2}s^{-1}$ white light conditions for ~20 hrs so cells were in mid log phase. A heterotrophic control culture containing no carbon source was used to confirm the lack of light in this condition. Another negative control with no inoculum in mixotrophic conditions was also included.

1 mL samples were taken at various time-points and cell density measured as OD_{750} in a spectrophotometer (UNICAM) to obtain a growth curve. At 12, 24, 48 and 120 hours, 14 mL samples were taken. 1 mL was used for OD_{750} measurement. 3 mL

for Nile Red assay to determine TAG accumulation, and 10 mL cell pellet for fatty-acid-methyl-ester (FAME) extraction to analyse chain length. Supernatant from the 10 mL sample was used to measure pH using a calibrated pH meter.

2.4.2 Comparing multiple carbon and nitrogen sources

For preparation of media components and carbon and nitrogen sources, all chemicals were ordered from Fisher or Sigma. All solutions were autoclaved for sterilisation, with carbon and nitrogen sources sterilised in separate bottles before addition to other media ingredients.

All cultures were grown mixotrophically (~150 µmol photons $m^{-2}s^{-1}$ fluorescent white light) as 200 mL volume in 500 mL flasks inoculated at 1 × 10^6 cells/ml with a mid- to late-log phase seed culture (grown in standard TAP media), continuous shaking at 120 rpm and 25 °C constant temperature (New Brunswick Scientific Innova™ 4430, 4340 or 44 incubator shaker). Cultures were harvested after 48 hours of growth by centrifugation at 5405 g (Sorvall® Evolution RC), removal of supernatant, snap-freezing of cell pellets in liquid nitrogen and freeze drying (Edwards Modulyo EF4) followed by longer term storage at -80 °C (up to 4 weeks). Before harvest a 10 mL sample was taken for cell count (Hawksley Improved Neubauer BS.748 Depth 0.1 mm, 1/400 mm^2; viewed under 40 x A-Plan objective lens using ZEISS AXIO Lab.A1 light microscope) and assay of neutral lipid amount by Nile Red staining. The pH of the culture supernatant was measured using pH indicator paper (FisherScientific). A known amount of the freeze dried sample was used for direct transesterification of lipids to fatty-acid-methyl-esters (FAMEs), which were analysed by gas chromatography (GC) with internal standards to reveal distribution of fatty-acid chain length. For each conditions there was biological replicates of n=3 or n=4.

2.5 Lipid analysis of *Chlorella sorokiniana*

Lipid analysis was carried out on wet or freeze-dried biomass. For freeze-drying, algal cells were freeze-dried until no weight change occurred.

2.5.1 FAME-GC

Lipid chain length profile analysis by fatty-acid-methyl-ester (FAME) transesterification and gas chromatography (GC). All GC work took place at Rothamsted Research under the kind supervision of Dr Mary Hamilton and Dr Richard Smith.

2.5.1.1 One step preparation of FAME from liquid pellet

10 mL sample was spun down in falcon or universal tube at 700-900 g 10 min in a benchtop centrifuge (model) and the supernatant transferred to new tube apart from a small amount (~ 1 mL). Pellet was re-suspended in remaining supernatant and transferred to 1.5 mL Eppendorf and spun down at 5000-7000 rpm for 2-5 min, supernatant discarded and pellet snap frozen in liquid nitrogen then transferred to storage at -80 °C. Alternately, supernatant was entirely removed from original tube, snap-frozen and stored at -80 without transfer to Eppendorf.

Cell pellets were transported to Rothamsted on dry ice. Pellets were transferred to glass methylation vials directly or using minimum amount of water (max 50 µl). 0.8 mL methylation mix (Methanol: Toluene: Dimethoxypropane: H_2SO_4; 33:14:2:1 by volume) followed by 0.8 mL heptane was added to sample using glass Pasteur pipette. Samples were baked at 80°C for 1-2 hrs, tubes vortexed briefly (2-5 s) and placed at -20°C or 4°C for ~15 minutes to allow phases to separate. Heptane layer (bottom layer) was transferred to GC vial and concentrated under nitrogen gas and placed in 200 µl insert if necessary. Samples were analysed on the GC (Agilent J&W DB-23-30M) with flame-ionisation detection (Programme: 50 °C 1 min hold, 40 °C/min to 190 °C, 1 °C/min to 220 °C. Hold 5 min, no split).

2.5.1.2 One step preparation of FAME from Freeze-dried pellet

A known amount of the freeze dried sample (equivalent to 5×10^7 cells; generally between 2-50 mg) was weighed out in a 3 ml glass methylation vial. For the majority of the high concentration carbon conditions, FAME analysis was not completed due to lack of cell growth. For the remaining samples, 1 ml of methylation

mix (Methanol: Toluene: Dimethyloxypropane: Sulphuric Acid at 33:14:3:1 by volume) containing internal lipid standards of C15:0 and C23:0 at 12.5 ng/µl was added to each freeze dried sample, vortexed briefly, then incubated for 90 min at 85 °C in a heat block. After letting samples cool to room temperature, 1 ml of 1 % w/v NaCl and 1 ml heptane was added to each vial and vortexed briefly before centrifuging for 2 min at 500 g (GeneVac EZ-2 series). 300 µl of the top heptane phase was transferred to a glass GC vial with insert and evaporated under nitrogen gas at 38 °C (temperature to aid evaporation). Samples were re-suspended in 70-1100 µl (depending on exact mg freeze dried sample weighed out, with 10 mg corresponding to a resuspension volume of 300 µl) heptane with 0.01% butylated hydroxytoluene (BHT) as an antioxidant. Vials were then flushed with nitrogen gas before immediately screwing on cap to minimise oxidation of sample, and were stored at -20 °C.

Samples were analysed by gas chromatography (Agilent) using an Agilent J&W DB-23-30M column with a 1 µl injection, split ratio of 5:1 and 7.5 ml/min flow rate. Oven programme was 150 °C start, hold 2 min, 10 °C/min to 250 °C, hold 5 min. Detection was with flame ionisation detector (GC-FID) at 260 °C, 30 ml/min H_2 flow and 400 ml/min air flow.

2.5.1.3 GC Data analysis

Peaks on the chromatogram were assigned and integrated using Agilent Chemstation software with a calibration table from a 37-FAME standard component mix (Supelco), C16:2n6 standard (Larodan) and C16:4n3 standard (Cambridge Bioscience). Peak data was exported and further analysis and graph generation was completed using Microsoft Excel 2016. Peaks were either expressed as mol % ratios, or normalised for exact lipid mass and cell number using the known amount of internal C15:0 fatty-acid standard and known cell number corresponding to biomass of sample used for FAME analysis.

2.5.2 Nile Red

Nile Red is a fluorescent dye which binds intracellular lipids and other hydrophobic compounds, resulting in a shift in emission wavelength when bound. For samples of live *C. sorokiniana* culture supplemented with 10 % DMSO, binding of Nile Red to lipids results in a shift of the broad emission peak from 653 nm to 586 nm (Vonlanthen, 2013).

The following protocol was used for Nile Red staining, which is optimised for *C. sorokiniana* (Vonlanthen, 2013): 330 µl DMSO was added to 3 ml algal cells in a 4 ml cuvette (final concentration 10%). 22 µl of a 0.3 mg/ml Nile Red (Thermo) solution in acetone was added to the sample (final concentration of 2 µg/ml) and incubated for 2.5 min in the dark. Fluorescence was measured in a fluorometer (Perkin Elmer LS55) with scanning emission wavelength of 530-750 nm and excitation at 510 nm. If the reading went above the maximum detection limit of the machine, the sample was diluted with TAP medium by a known factor. Data was exported using the WinLab software and processed using Microsoft Excel 2016. The fluorescence reading (arbitrary units) were normalised to cell number by dividing by number of cells in the cuvette.

2.5.3 Total Lipids

2.5.3.1 Total Lipid extraction

Two different methods were used for lipid extraction as discussed in the results section. If the concentration of samples was required, this was carried out with a GeneVac EZ-2 machine using 30 °C/low boiling point programme under nitrogen gas/purge.

For the first method (Vonlanthen, 2013), 50 mg lyophilised cell pellet was weighed into a 4 mL cryovial (Nunc/Sigma). 200 mg acid washed glass beads and a ceramic ball were added to each tube along with 1 mL chloroform: methanol (2:1 v/v). Cells were lysed by vortexing tubes for 1 min, incubating for 15 min and vortexing for another 30 s at room temperature. 0.3 mL MeOH and 0.3 mL dH_2O were added and vortexed for 1 min. Phase layers were separated by centrifugation at 2500

g for 3 min. A glass Pasteur pipette was used to carefully extract (avoiding disturbance of debris layer between phases) the bottom layer into a clean glass vial which was used directly for further analysis or stored at -20 °C (short term <1 week) or -80 °C (long term > 1 week).

The second method (Axelsson and Gentili, 2014) was as follows: 50 mg lyophilized cell pellet was homogenised by grinding with a pestle and mortar in liquid nitrogen and transferred to a 30 mL lidded glass centrifuge tube (Kimble) and subsequently adding 8 mL chloroform:methanol (2:1 v/v). Biomass was resuspended by shaking vigorously for a few seconds until biomass evenly dispersed according to eye before adding 2 ml 0.73% NaCl. Tube was spun for 2 min at 350 g and lower phase recovered using a glass Pasteur pipette into a clean glass vial with care to avoid disturbance of the biomass debris layer between the phases.

2.5.3.2 TLC

TLC was performed on total lipid extractions which had been dried under nitrogen gas. A TLC tank was prepared for 1 hour with a solvent mix of hexane : diethyl ether : acetic acid (75:25:1). The TLC plate was activated by baking at 80 °C for 1 hour. Total lipid extracts were dried under nitrogen gas and re-suspended in 80 µl $CHCl_3$ using a Hamilton Syringe, which was subsequently loaded onto the TLC plate. Additionally, 20 µl and 10 µl 17:0 TAG standard was loaded. The plate was dried under nitrogen gas and incubated in a sealed solvent tank for 1 hour then dried for 10 min on blue roll in a fume hood. The plate was visualised by spraying with 0.05% primuline in acetone:water 4:1 and observing under UV light. Spots were marked with a pencil and scraped off using a sharp tool into a GC vial, covered with $CHCl_3$ and stored at -20 °C until further analysis.

2.6 Bioinformatics analysis of the published PBCV-1 chlorovirus transcriptome

Tools used include NCBI Blast (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and the Sequence Read Archives (SRA): https://www.ncbi.nlm.nih.gov/sra. Sequences were visualised and annotated using Benchling https://benchling.com/.

2.7 Nuclear transformation of *Chlorella sorokiniana*

2.7.1 Transformation of *C. sorokiniana* using *Agrobacterium tumefaciens*

The procedure detailed here was previously developed in the lab (Hiegle, 2014) but further optimisations and modifications are described in the results section. Wild-type *C. sorokiniana* was inoculated from a TAP agar plate stock into 50-100 ml TAP into a 100 – 250 ml shake flask and grown until early log phase (~20-35 hours). 20 million cells (concentrating cells if necessary) were then plated out on TAP agar containing Acetosyringone at 150 µM and incubated for one day. At the same time the *A. tumefaciens* strain EHA105 containing the desired binary vector was cultivated by growing from a glycerol stock in 6 ml YENB media containing 50-100 µg/ml kanamycin in a 30 mL sterilin for 1 day, shaking ~300 rpm at 28 °C. 200 µl of ~OD_{600} 1 *Agrobacterium* culture which has been supplemented with 150 µM acetosyringone 2 hours prior, is spread onto the *C. sorokiniana* plate and mixed thoroughly using a plate spreader and left in dark in growth chamber at 25 °C for 2 days. *C. sorokiniana* cells are then recovered by harvesting cells from plate by resuspending and transferring to 30 mL sterilin in 25 mL TP medium supplemented with carbenicillin (250 µg/ml) and incubating overnight in the light at 25 °C with gentle shaking.

Putative transformants were then selected either directly on solid medium or with an additional selective liquid medium enrichment step. For direct solid selection, ~90 million cells were plated out using soft agar overlay onto TAP agar plates containing 300 µg/ml G418, 500 µg/ml cefotaxime and 500 µg/ml carbenicillin and incubated in the light at 25 °C until single colonies appear (~9 days). For liquid selection enrichment: 130 million cells in maximum 1 ml volume (cells concentrated by centrifugation at 900 g if necessary) are added to 50 mL TAP supplemented with 50 µg/ml G418 and 250 µg/ml carbenicillin and incubated in the light with shaking at 120 rpm, 25 °C, for ~5 days or until cultures began to green. At this point, 0.5 mL selective culture is plated using soft agar overlay on TAP agar plates supplemented with 350-450 µg/ml G418, 500 µg/ml carbenicillin and 500 µg/ml cefotaxime and incubated in the light for 25 °C until single colonies appear (~9 days).

Negative control transformations including *Agrobacterium* with no binary plasmid, and no *Agrobacterium* are also included to identify likelihood of spontaneous colonies appearing on putative transformant plates.

2.7.2 Electroporation

A 1L mid-late log phase cell culture (1.5×10^7 cells/ml) grown mixotrophically in TAP (25 °C, 120 rpm, ~150 µmol photons $m^{-2}s^{-1}$) was harvested at 4 °C, 2500 g, 5 min in sterile 50 mL falcon tubes and each pellet was washed once with 40 mL TAP + sucrose (40 mM) by pipetting up and down 20 times and discarding supernatant. Tubes were tapped dry on paper towels then pellet resuspended with approximately 250 µl TAP + sucrose (40mM) and pooled to a 1.5 ml Eppendorf tube; volume of TAP added was adjusted to achieve a total of 1 mL cell suspension from the original 1 L culture, resulting in a concentration of 1 $x10^{10}$ cells/mL. Cells were kept on ice and an aliquot was diluted with TAP + sucrose (40 mM) to concentrations of 1×10^{10}, 1×10^8 and 1×10^6 cells/mL. 300 µL cells at each concentration were added to a 0.2 cm sterile electroporation cuvette (Nepa EC-002 40-400 µL) with 2 µg linear or circularized plasmid. The ratio of cell number to amount of DNA (µg) was 1.5×10^9, 1.5×10^7 and 1.5×10^5 for the different cell concentrations. Negative controls containing no DNA were also included by substituting DNA for sterile water.

The mixture was electroporated using the following settings: exponential decay, 25 µF capacitance, 2000 V (BioRad GenePulser XCell). The resistance was not able to be set so was left at infinity. Electroporated cells were immediately transferred to 10 mL TAP sucrose (40 mM) in 15 mL falcons at room temperature. Recovery cultures were subsequently incubated overnight (~14 hrs) horizontally in identical conditions to previously except covered by 4 sheets of white paper towel to reduce light intensity. Following incubation, cells were harvested by centrifugation at 2000 g for 5 min followed by 2500 g for 2 min, supernatant discarded and pellet resuspended in 200 µL TAP media. Re-suspended cells were plated on TAP agar (2 %) plates containing either G418 (Gibco Geneticin) at 100 µg/mL or no antibiotic as a positive control. Plates were incubated 25 °C, ~150 µmol photons $m^{-2}s^{-1}$ continuous white light. Plates were examined firstly after 4 days. The number of colonies on each

plate was monitored on day 8 and day 23 by taking pictures of plates with colonies and estimating the amount using OpenCFU software version 3.9.0 (Geissmann, 2013) with settings of Regular Threshold 2 or 3 and a minimum diameter of 1.

2.8 Nuclear transformation of *C. reinhardtii* using glass beads

The method used was based on (Kindle, 1998) as used in (Vonlanthen, 2013). A starter culture was used to inoculate 500 mL TAP at 1-2 % (v/v) inoculum and grown to mid-log phase corresponding to approximately 2×10^6 cells/ml. After harvesting the algae by centrifugation at 3800 g, room temperature (Benchtop Megafuge) the cells were resuspended in fresh TAP medium to a concentration of 2×10^8 cells/ml. 300 µl cells were transferred into a glass test tube containing 300 mg acid-washed glass beads (Sigma) and 1 µg Level 1 plasmid linearised with *NheI* restriction enzyme which cuts once in the vector backbone. Test tubes were agitated using a vortex (VortexGenie) for 15 sec set at top speed before transferring the cells into a 30 mL sterilin tube containing 10 mL fresh TAP medium and recovering overnight in constant light with shaking at 120 rpm. For selection the recovered cells were harvested by centrifugation at 3800 g at room temperature for 5 minutes and cell pellets resuspended in 3.5 mL soft TAP agar (0.5 % w/v) before plating as on overlay onto solid agar plates containing 5-10 µg/ml of the selective antibiotic Zeocin™ (Invitrogen). Plates were incubated in constant light at 25 °C until colonies appeared, after which they were propagated by re-streaking on selective TAP agar.

CHAPTER 3 CHARACTERISATION OF *CHLORELLA SOROKINIANA* GROWTH AND LIPID PROFILE IN DIFFERENT CONDITIONS

3.1 Introduction

Manipulation of growth conditions is a common strategy used to modify the lipid content and fatty-acid profile of microalgae (Hu et al., 2008). *C. sorokiniana* UTEX 1230 has received much interest because of its fast growth rate (doubling times of as low as 4 hours, it is able to reach stationary phase after a day and a half of cultivation in shake flasks) and ability to accumulate high levels of TAGs, especially when an exogenous carbon source is supplied. For example, one report found that UTEX 1230 accumulates 3.8-fold more lipid in heterotrophy than in phototrophy (0.53 g/L compared to 0.14 g/L), twice as much total biomass and twice as much TAG proportionally across total lipids (Rosenberg et al., 2014). Another report using UTEX 1230 in outdoor conditions found that biomass yield was higher in mixotrophy than in heterotrophy (Lee et al., 1996), however this does not agree with the data from Rosenberg *et al.* who found that although mixotrophic conditions performed better than phototrophic they still did not reach heterotrophic productivities (Rosenberg et al., 2014). For *C. sorokiniana* UTEX 1602 mixotrophic conditions were also reported to yield higher biomass and lipid contents (Li et al., 2014). These reports all used glucose as the exogenous carbon source, and highlight the large variability in lipid studies which are likely due to small differences in the environmental conditions or growth media.

It is consistent though that comparatively, *C. sorokiniana* UTEX 1230 does not perform as well in phototrophic conditions. Surprisingly, one study found only minor change in lipid content when UTEX 1230 was grown on low nitrogen medium, the only one out of five *Chlorella* species they surveyed (Illman et al., 2000). The UTEX 1230 cells did exhibit three-fold reduced growth rate for when grown in low nitrogen medium but examination of the growth curve appears to show the UTEX 1230 strain

declining in cell number at the end of the experiment in contrast to the other species (Illman et al., 2000). It could be that the decline of the culture prevented proper lipid accumulation. Another study comparing lipid productivity across various *Chlorella* species found that strains supposedly identical to UTEX 1230 were among the lowest values (Přibyl et al., 2012). For both these reports, the cells were cultivated under phototrophic conditions (i.e. using only carbon dioxide as the sole carbon source) which evidently is not optimal for UTEX 1230. Supply of exogenous carbon is required to maximise productivity from this strain.

This high productivity makes UTEX 1230 an attractive industrial strain. One closely related species, *C. sorokiniana* DOE 1412, was the subject of extensive characterisation in a US Department of Energy study who concluded that it had promise for biofuel applications (Lammers et al., 2017; Neofotis et al., 2016). When grown on anaerobic digester effluent from cattle manure digestion, UTEX 1230 accumulated starch and proteins rather than lipids (max. 35 % of dry weight) whereas on a basal medium (BBM) the percentage of lipid dry weight reached nearly 60 % (Kobayashi et al., 2013). Inoculum density has also been studied for its impact upon the productivity and fatty acid profile of microalgae, including *C. sorokiniana* where larger inoculums were found to have better productivity (Lu et al., 2012) and had changes in phospholipid species abundances and chain lengths. (Lu et al., 2013)

The high productivity, fast growth rate and robustness to high light and temperature make UTEX 1230 suited to outdoor cultivation. The production of high value omega-3 fatty acids such as EPA and DHA in such a productive strain is an attractive biotechnological prospect. The fatty acid chain length profile of *C. sorokiniana* UTEX 1230 has been previously characterised and naturally produces chains of up to a maximum of 18 carbons including C18:3n3 (ALA) (Kobayashi et al., 2013; Lang et al., 2011; Petkov and Garcia, 2007; Rosenberg et al., 2014; Tang et al., 2016b; Vonlanthen, 2013). This is the precursor fatty-acid to subsequent elongation and desaturation in the omega-3 fatty-acid synthesis pathway (**Figure 1-10**). Therefore, if LC-PUFAs were to be produced in this strain, transgene expression is necessary, including genes encoding heterologous elongases and desaturases. A

total of five heterologous genes would be needed to complete the omega-3 pathway to include DHA (**Figure 1-10**).

3.1.1 Aims and brief experimental layout

As described above and in the introduction section 1.4, there are many factors which influence growth and lipid profile in microalgae, ranging from basic culturing techniques to abiotic stressors and the specific growth media used. This chapter aims to explore some factors affecting both overall growth and lipid accumulation (**Figure 3-1**), with a focus on identifying a condition suitable for preferential accumulation of C18:3n3 (ALA), the precursor fatty-acid to the omega-3 pathway. If a transgenic strain of UTEX 1230 becomes available containing a completed omega-3 pathway, then increasing the amount of this C18:3n3 substrate is desirable.

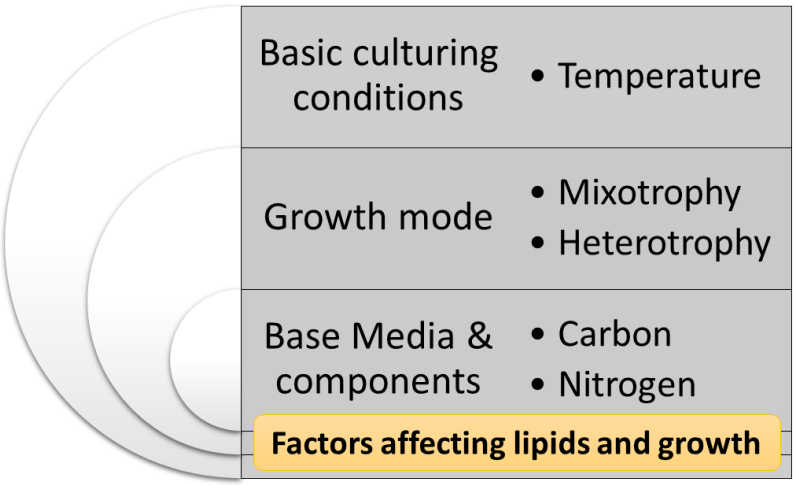


Figure 3-1 Factors influencing lipids and growth which are explored in this chapter.

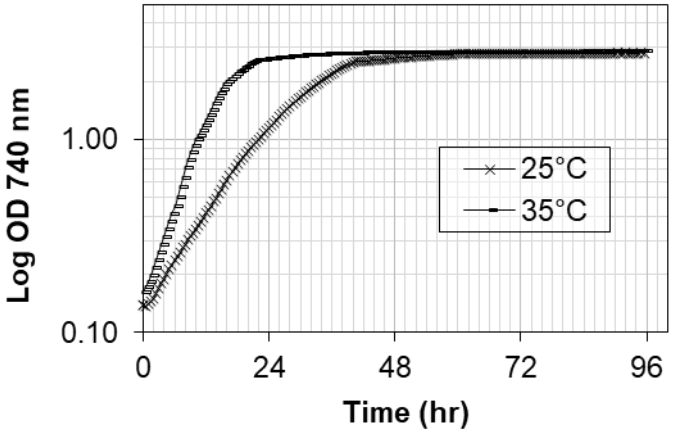
This chapter starts with examining the growth of UTEX 1230 under different basic culturing conditions such as temperature, trophic mode and variation of the base media (section 3.2). After examining some lipid methodology (section 3.3) the influence of these same factors on the FAME profile of UTEX 1230 is examined (section 3.4). The chapter finishes with a larger survey of carbon and nitrogen sources, which can be used as stressors for lipid accumulation via deprivation or in excess (section 3.6). In addition to effects on cell growth and FAME profile (focusing on C18:3n3 accumulation) this latter experiment also describes preliminary data on

the neutral lipid class (an indication of TAG level) because this form of lipid is industrially desirable. The consideration of the lipid profile in the context of algae growth is also important to gain insights into overall productivity.

3.2 Assessing the growth of *Chlorella sorokiniana* UTEX 1230 in different conditions

3.2.1 *Chlorella sorokiniana* exhibits increased growth rates at higher temperatures

Using the Algem photobioreactor, which can incubate two separate 1 L flasks (400-500 mL working volume) in controlled conditions, a comparison of *C. sorokiniana* mixotrophic growth at two different temperatures of 25 °C and 35 °C was carried out in TAP medium with a light supply of 200 µmol photons $m^{-2}s^{-1}$ PAR (Photosynthetically Active Radiation: wavelength 400-700 nm). The two 500 ml cultures were inoculated using 1% liquid inoculum from a seed culture. Both Algem cultures reached the same OD for stationary phase but it took approximately 12 hours longer for the lower temperature to reach stationary phase (**Figure 3-2**).



Temp	Final manual OD 750	Final cell count (cells/ml)
25 °C	5.70	99×10^6
35 °C	4.75	121×10^6

Figure 3-2 Growth of *Chlorella sorokiniana* UTEX 1230 at two temperatures in the Algem photobioreactor. Cell density was measured automatically every hour in cultures at 25 °C (solid diamonds) or 35 °C (empty diamonds) in the Algem photobioreactor using TAP medium and illumination of 200 µmol photons $m^{-2}s^{-1}$ PAR.

The doubling times were also different with the higher temperature being twice as fast at 3.5 hours compared to 7 hours, as calculated from the log phase gradient on the graph. A manual OD_{750} reading and a manual cell count were taken at the end of the experiment at 96 hours. For the two temperatures, the Algem appears to suggest they have same final OD (**Figure 3-2**), looking at the manual OD values the 25 °C culture is 1.2 times higher. However, the final cell density was 1.2 times lower at 25 °C. Discrepancies between optical density and cell density measurements can occur when physical characteristics such as cell size are different between samples. The temperature could be causing a morphological change in the cells that could be interesting to explore further using microscopy.

3.2.2 *Chlorella sorokiniana* exhibits changes in growth rate and stationary phase cell density in different growth modes and media.

Microalgae are versatile organisms that can grow in a variety of environments and substrates. To explore the effect this has on *C. sorokiniana*, the growth modes of

mixotrophy (organic carbon source and light source available) and heterotrophy (organic carbon source but no light source) were compared. Additionally, two different commonly used base growth media were examined. These experiments are split into two sections: Mixotrophy and heterotrophy were first explored in a preliminary experiment using the controlled conditions of the Algem photobioreactor (section 3.2.2.1). Then, the exploration of growth mode is extended further in a shake flask experiment where two different base media are used (section 3.2.2.2). The growth media chosen are TAP and BBM. Both are commonly used for growth of green algae.

3.2.2.1 Mixotrophy and Heterotrophy in TAP comparison in the Algem photobioreactor

Mixotrophic and heterotrophic growth modes were compared in the Algem photobioreactor in TAP medium at 27 °C by growing under 100 constant white PAR (mixotrophic) or no light (heterotrophic) (**Figure 3-3**).

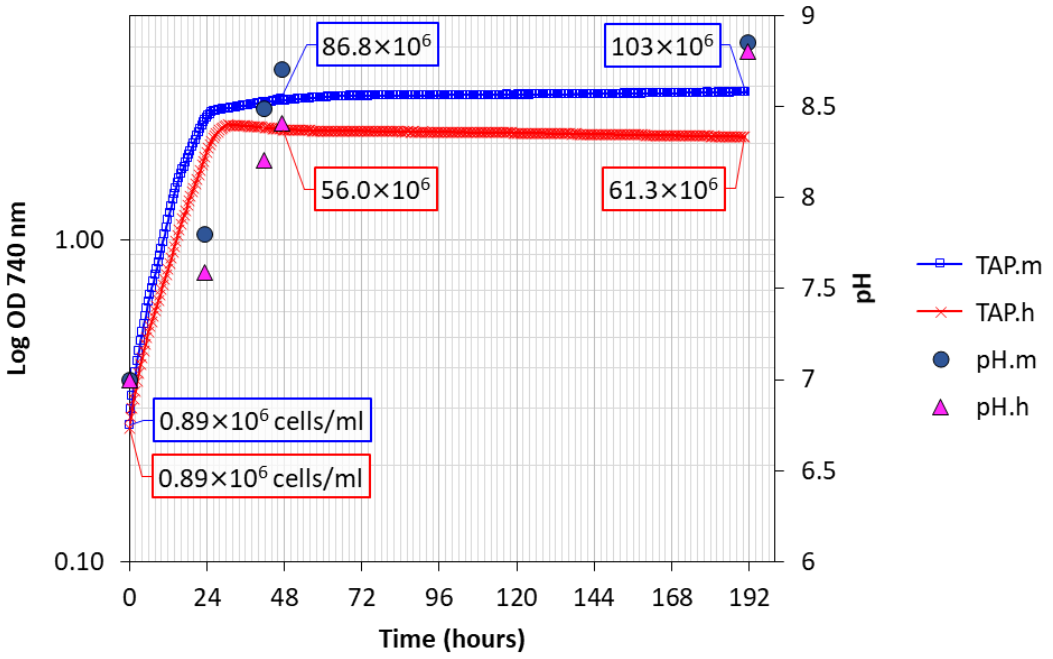


Figure 3-3 Comparison of mixotrophic and heterotrophic growth in the Algem photobioreactor. Boxed numbers are manually measured cells/ml. A final manual OD_{750} reading was 1.77 for heterotrophy and 4.93 for mixotrophy.

The mixotrophic culture grew with a slightly faster growth rate as seen by the steeper gradient on the graph (**Figure 3-3**). The mixotrophic culture also supported a higher cell density. The mixotrophic culture reached stationary phase at 24 hours, slightly earlier than the heterotrophic culture at about 30 hours.

The pH showed a steady increase over time for both mixotrophy and heterotrophy from pH 7 to a final pH of 8.75, with the sharpest rise occurring whilst the cells are in the exponential phase (**Figure 3-3**). This rise in pH is considered further in the discussion. At the final timepoint a manual OD 750 was taken which was 1.77 for heterotrophy and 4.93 for mixotrophy. The OD 750 increase is over 2.7-fold for mixotrophy compared to heterotrophy, whereas for cell number mixotrophy is only 1.6 time higher than heterotrophy. This disproportionate increase in OD_{750} compared to cell density could be due to biomass and the size of the individual cells.

3.2.2.2 Mixotrophy, heterotrophy and two different base media in a shake flask experiment

C. sorokiniana growth was monitored by measuring OD_{750} at fixed time points for both mixotrophy and heterotrophy. The two growth media, TAP and BBM-Glucose (BBMG), were compared, which contain different amounts of carbon and nitrogen (**Table 3-1**). Heterotrophic control cultures containing no carbon source (TP and BBM media) was used to confirm the lack of growth without light in this condition. Growth on BBMG medium is of interest since one study found increased TAGs and much higher growth and cell density for UTEX 1230 when grown on BBM supplemented with glucose in heterotrophic conditions (Rosenberg et al., 2014). BBM glucose supplementation amount, inoculation density (~10^6 cells/ml) and temperature (25 °C) are matched to the conditions used in the study. The actual inoculation density was 0.6×10^6 for BBMG and 1×10^6 for TAP because of the different growth rates of the seed cultures, which were grown for the same length of time before being used.

Media	C source	N source
BBMG	Glucose 55.5 mM (10 g/L)	Nitrate 2.94 mM
BBM	None added	Nitrate 2.94 mM
TAP	Acetate 17.4 mM	Ammonium 7.48 mM
TP	None added	Ammonium 7.48 mM

Table 3-1 Carbon and Nitrogen sources in BBMG and TAP media. Concentration of Carbon and Nitrogen sources in Bold's Basal Media (BBM) supplemented with glucose (BBMG) and Tris-Acetate-Phosphate (TAP) media.

Mixotrophic conditions in both media supported faster growth than in heterotrophic conditions, with both media reaching similar final OD 750 (**Figure 3-4**). As calculated from the log-phase of the graph, the mixotrophic TAP doubling time is ~5 hours, which was faster than the ~7 hours achieved in the Algem photobioreactor under equivalent conditions (**Figure 3-2**). Under heterotrophy ('.h'), BBMG supported higher final OD_{750} after 120 hours than TAP, despite slower initial growth. The higher final OD_{750} could be due to the increased carbon source available than in TAP medium (**Table 3-1**). Additionally, the BBMG.h culture appears to not have reached stationary phase by the end of the experiment suggesting that this medium could support a higher density of cells overall. True heterotrophic conditions were confirmed by lack of growth of the negative control culture grown without carbon source (TP.h), but the growth of BBM.h (BBM with no supplemental glucose) shows that there must be some other form of carbon source available to the algae in this sample.

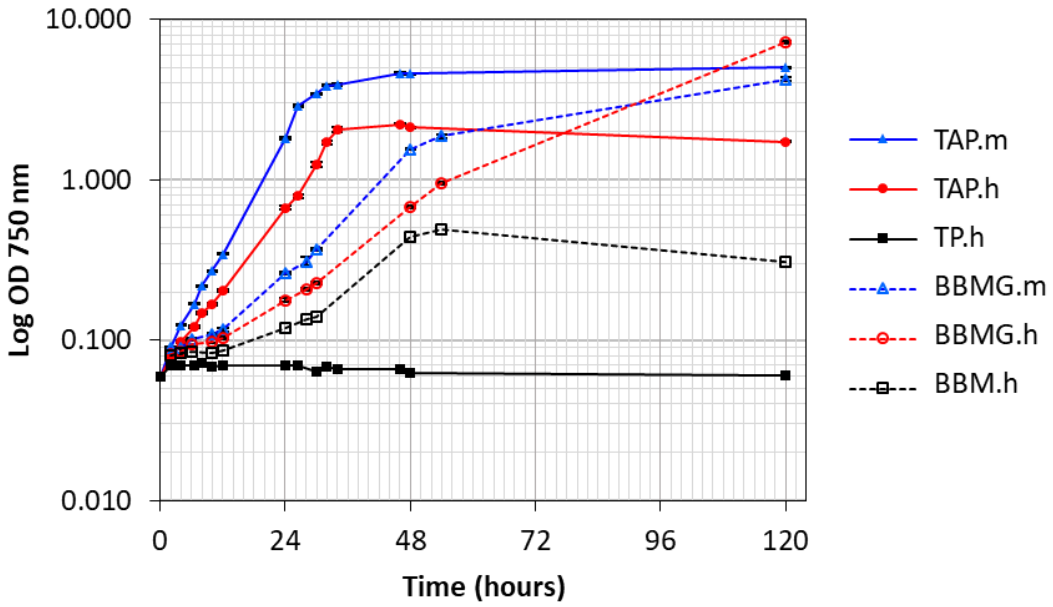


Figure 3-4 Growth of wild-type *C. sorokiniana* in two different growth modes and media. Cells were grown in shake flasks in TAP (solid lines and markers) and BBMG (dotted lines and empty markers) in mixotrophic (triangles, "m") or heterotrophic (circles, "h") conditions. Two heterotrophic conditions with no carbon source were included for negative controls (TP and BBM, square markers) to test any residual carbon in the two different base media. The temperature was 25 °C; and illumination by a white fluorescent light strip was ~150 µmol photons $m^{-2}s^{-1}$. Standard error bars are shown (n=4).

The pH was also monitored (**Figure 3-5** and showed that BBM/BBMG medium is more acidic than TP/TAP, with the former starting at pH 5.5-6 compared to 7-7.5 in the latter. Growth in both media and growth modes increased pH compared to the non-inoculated negative control, reaching a maximum of 8.75 for TAP.m and TAP.h, 7.5 for BBM.m and 6.5 for BBM.h. These increases in pH are broadly aligned with the growth curves of cells, as seen in the Algem experiment (**Figure 3-3**). However, it is interesting to note the rise in the non-inoculated control in BBM from 5.75 to 6.25, suggesting this media could be undergoing chemical changes, or that this change is caused by the small amount of cell growth in this condition (**Figure 3-4**).

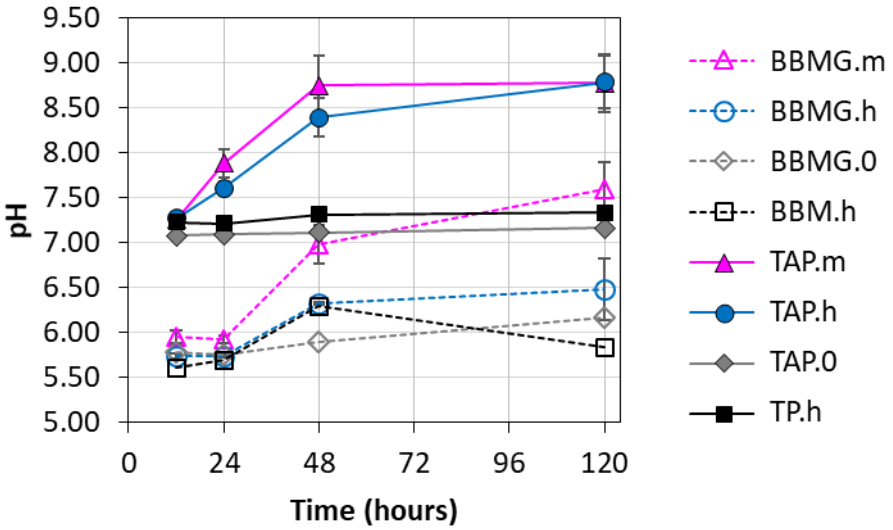


Figure 3-5 Changes in pH in mixotrophic and heterotrophic cultures of *C. sorokiniana* in two different media. Mixotrophy (m), heterotrophy (h), no cells (0). Heterotrophic controls with no additional carbon source (BBM, TP) were also included. Standard error bars (n=4).

3.2.3 Summary of observations for basic culturing conditions

Data was collected to explore the effect of temperature, growth mode and base media on the growth and FAME profile of *C. sorokiniana*.

In TAP medium, mixotrophy supports faster growth and higher final OD_{750} than heterotrophy. For BBMG, mixotrophy supports faster growth until ~48-72 hours at which point it begins to approach stationary phase, whereas the heterotrophic condition continues to increase in OD_{750} until the end of the experiment at 120 hours. For both BBMG and TAP, heterotrophic growth does not appear to have reached stationary phase by the end of the experiment, especially for BBMG. This suggests that heterotrophic conditions may be more productive in an industrial setting where maximal accumulation of product is desired.

The final OD_{750} differences in the BBMG.h and BBMG.m conditions are interesting because it suggests there are different limiting factors to growth. For example, shading caused by increased biomass would not be limiting for BBMG.h but could be for the mixotrophic growth if photosynthesis is contributing to cell growth

in this condition. To properly assess which nutrients becomes limiting, analysis of the culture metabolites or spent media by a technique such as HPLC would be needed, which could be a useful future experiment.

The pH increases with cell growth in all media and growth conditions, but BBM media overall has a lower pH. The pH changes broadly follow the growth curves of the microalgae in all conditions. Therefore, pH change could be used as a sign of active metabolism during microalgal growth. In TAP, the rise in pH can be explained because as acetate is assimilated by the algae, it reduces the amount of acidic component of the Tris:acetate buffer system in the media. For BBMG medium the rise in pH is probably due to nitrate uptake which is associated with a rise in pH in algal cultures (Scherholz and Curtis, 2013; Tan and Adebusuyi, 2016).

3.3 Lipid Analysis Methods

The experiments in the subsequent section concern the exploration of lipid content and profile of cells to put this into context with the above growth studies. Lipid analysis can be a time-consuming part of algal analyses and a source of unwanted variation (Cavonius et al., 2014). Therefore, some tests and verifications were carried out relating to the methods used.

3.3.1 A comparison of two methods of total lipid extraction

Adequate lipid extraction is important for analysis of promising industrial microalgal strains or characterisation of putative transformants but the process can be time consuming and can be highly variable between different studies rendering different datasets incomparable (Cavonius et al., 2014). To research an appropriate analysis method, two solvent-based methods were compared using 50 mg of freeze-dried cell pellets from two pooled 120 hour cultures grown heterotrophically in BBMG. The cultures had a mean cell density of 75.5×10^6 cells/ml and mean OD_{750} of 6.20.

Protocol A involves the use of glass beads and a ceramic ball to aid mechanical cell lysis while vortexing and the solvent mix was chloroform:methanol (2:1 (v/v))

followed by phase separation using methanol and distilled water (1:1) (Vonlanthen, 2013) (**Table 3-2**). Protocol B involved manual homogenisation using a pestle and mortar in liquid nitrogen, and the solvent mix was the same except using a larger volume, followed by phase separation using 0.73 % NaCl (Axelsson and Gentili, 2014) (**Table 3-2**). Total lipid (mg) was measured gravimetrically after sample evaporation under nitrogen gas. Detailed protocols are described in Materials and Methods section 2.5.

Method	Reference	Physical Cell disruption	Solvent Mix	Phase Separation	Lipid extracted (mg)
A	(Vonlanthen, 2013)	Agitation in a tube with a ceramic ball and 200 mg acid washed glass beads using a vortex	Chloroform: Methanol 2:1 (v/v), 1 mL	0.3 mL Methanol, 0.3 mL distilled water	2.8 ± 1.3 (n=4)
B	(Axelsson and Gentili, 2014)	Homogenisation of cells in liquid nitrogen with pestle and mortar	Chloroform: Methanol 2:1 (v/v), 8 mL	2 mL 0.73 % Sodium Chloride (w/v)	3.1 ± 0.78 (n=3)

Table 3-2 Comparison of two different solvent-based lipid extraction methods used on 50 mg freeze-dried biomass of *Chlorella sorokiniana* UTEX 1230. Extracted lipid is expressed as mean average with standard deviation.

Both lipid extraction methods gave similar amounts of total lipid (**Table 3-2**) corresponding to 5.6% and 6.2% of total biomass for methods A and B, respectively. This was expected since the solvent mixes consisted of the same chemicals, and confirms that the difference in physical cell disruption and phase separation steps only have a negligible effect. In order to assess if either method was biased towards certain TAG lipid species, the extracts were analysed further using thin-layer-chromatography (TLC, **Figure 3-6**). In this preliminary example, the resulting plate image has a clear large TAG spot for both methods towards the top of the plate, suggesting the sample is mainly TAG, although there are other lipids such as polar lipids and pigments visible at the bottom (**Figure 3-6**).

The TAG spots were scraped off the plate for analysis of carbon number and positions in the TAGs, which was carried out by Guillaume Menard at Rothamsted Research. The consistency of the profile across both methods shows that both the methods do not vary in extraction across chain length or “sn” position within the TAG molecules (**Figure 3-7**). Since both methods give the same result, method A was chosen because it is less labour intensive.

Figure 3-6. Separation of heterotrophically grown *C. sorokiniana* lipid extracts by thin-layer chromatography. Cells were grown in BBM + 10 g/L glucose, 25 °C, 120rpm, no light, for 120 hrs. Plate was visualised with primuline staining. The samples are total lipid extracts using method A (1) and B (2) respectively. TAG spots are circled. The standards are two different concentrations of a representative TAG molecule.

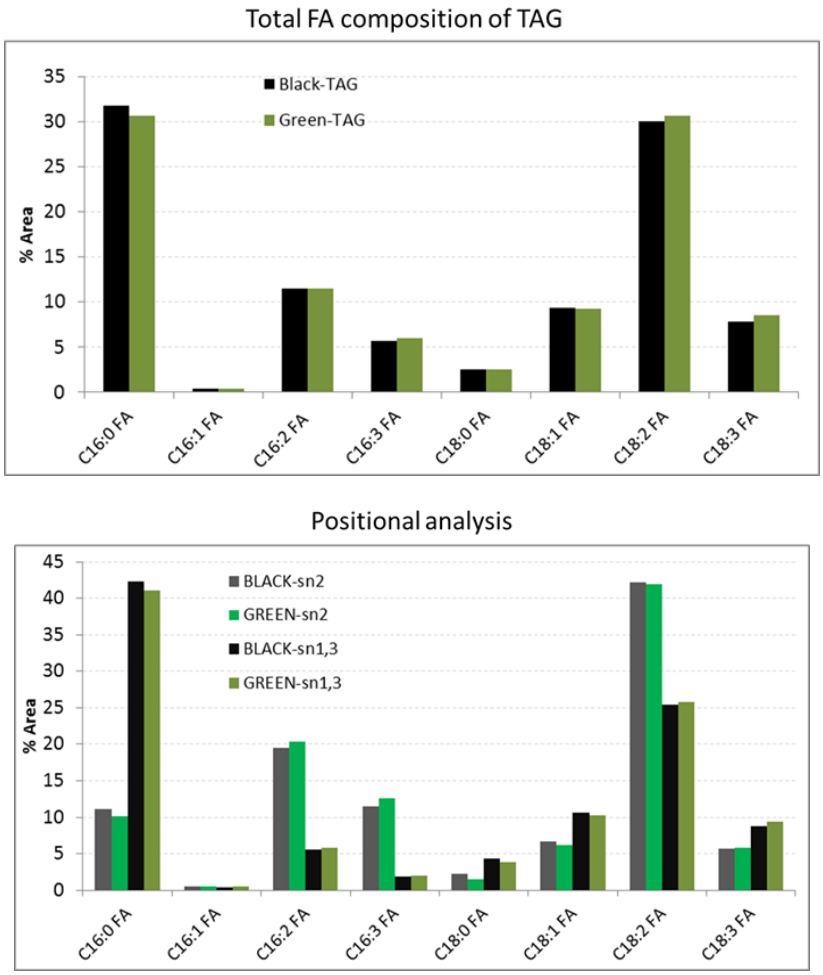


Figure 3-7 The mol% FAME profile and positional analysis of the TAG fraction of a heterotrophically grown *C. sorokiniana* culture in BBMG media for 120 hours. This analysis was performed by Guillaume Menard at Rothamsted Research. No major differences are seen between method A ("green") and B ("black") lipid extractions methods.

3.3.2 Presence of DMSO does not affect cell fluorescence

Nile Red is a lipophilic dye that is often used to assess the neutral lipid content of microalgae because in a hydrophobic environment the emission spectrum is blue-shifted from a peak at 660 to 586 nm (**Table 3-4**) (Greenspan and Fowler, 1985). In a protocol previously optimised for *C. sorokiniana* it was found to require a carrier solvent of DMSO to allow the dye to penetrate the tough cell wall (Vonlanthen, 2013). It is important to determine any background effect of this solvent on cell fluorescence so any Nile Red readings can be normalised if necessary. No effect of DMSO on cell fluorescence was found for three cultures at different cell concentrations after using

the recommended conditions of 10% DMSO and incubation time of 2.5 minutes (**Figure 3-8**). The peak visible at 680 nm is chlorophyll fluorescence which is unchanged by DMSO apart from potentially being marginally reduced. This could be due to lysis of cells since DMSO is toxic to *Chlorella sorokiniana* (Vonlanthen, 2013).

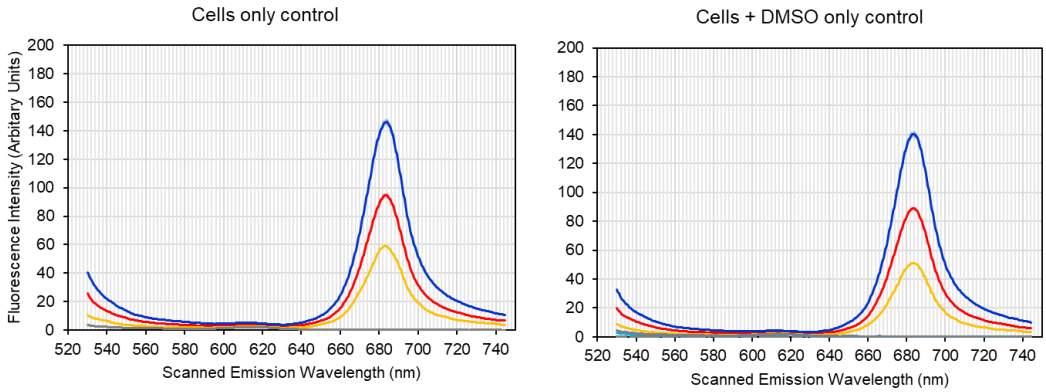


Figure 3-8 Influence of DMSO on fluorescence profile of wild-type *C. sorokiniana*. DMSO at 10% concentration and incubated for 2.5 minutes does not affect the fluorescence profile of wild-type C. *sorokiniana* cells. The three lines correspond to cultures of different cell densities. The culture was excited at 510 nm and fluorescence measured by scanning emission wavelength of 530-750 nm.

3.4 FAME profile over time is different under mixotrophy and heterotrophy across two different media

In this experiment the FAME profile complementary to the growth experiments presented in section 3.2 was measured. FAME samples were taken from the shake flask experiments (3.2.2.2) at 12, 24, 48 and 120 hours. FAME samples were also taken from the Algem growth experiment (3.2.2.1) at 17, 23, 42 and 47 hours. This data therefore covers the FAME profile over time for a variety of growth conditions including two trophic modes in two different media. The focus of this lipid analysis was to explore the FAME profile in particular in relation to ALA, the omega-3 pathway precursor fatty-acid, as it was desirable to explore which conditions favour the most of this fatty acid.

The data is presented in multiple different ways for easier comparison. The sections below compare the same data by changes in time (section 3.4.1), then mixotrophy and heterotrophy (section 3.4.2), different base media (section 3.4.3), and finally considering just the fatty acid of interest, ALA (section 3.4.4).

3.4.1 Change in FAME profile over time

The FAME profile varies greatly over time with the most pronounced changes seen in the BBMG medium for 16:1 which decreases over time and 18:2 which increases over time (**Figure 3-9**).

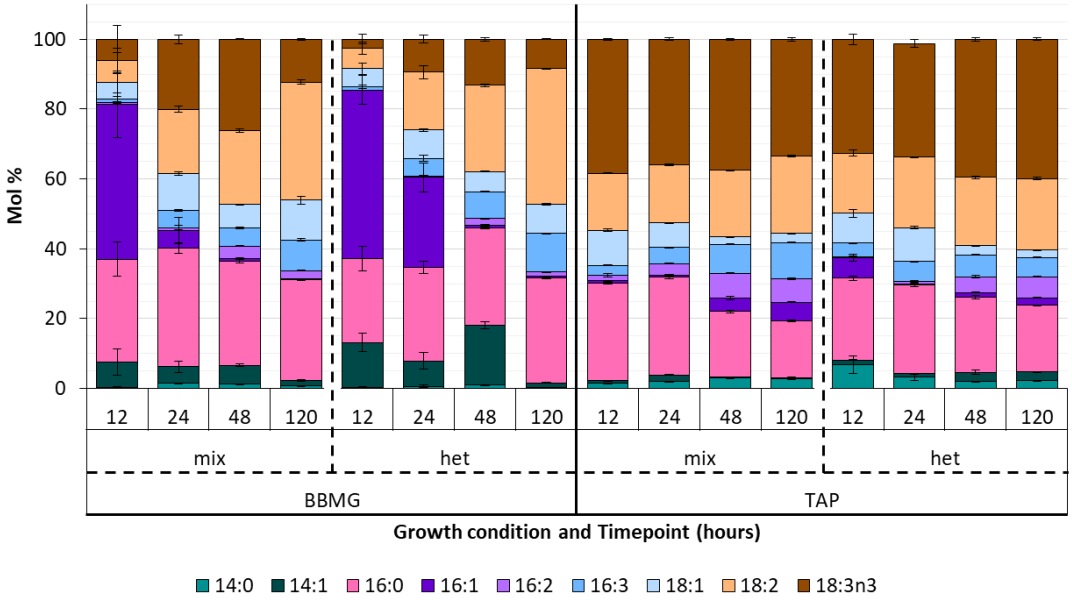


Figure 3-9 The FAME profile of *C. sorokiniana* over time in two different trophic modes and media. Standard error bars are included (n=4).

A difference in FAME profile is also seen from analysis of the Algem cultures when comparing the first timepoint (17 hr, log phase) to the other timepoints. At 17 hours there is increased 14:0, 14:1 and 16:1 but less 18:3n3, 18:2 and 16:0 (**Figure 3-10**).

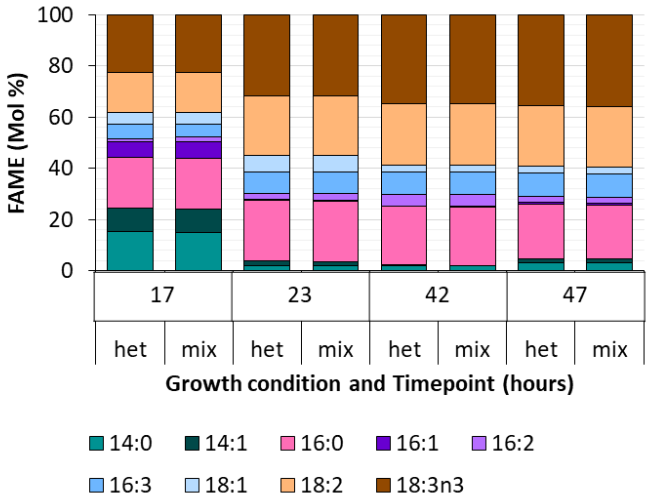


Figure 3-10 FAME profile of *Chlorella* sorokiniana UTEX 1230 over 2 days grown in two different growth modes in the Algem photobioreactor in TAP medium. The cells were grown heterotrophically (het) or mixotrophically (mix) by controlling the light level to be zero or 100 PAR respectively.

3.4.2 Comparing mixotrophy and heterotrophy

In the Algem experiment (TAP medium) there are minimal differences in FAME profile between mixotrophy and heterotrophy apart from a small presence of 16:2 at 17 hours under mixotrophy, but not in heterotrophy (**Figure 3-10**). This is also true for the TAP in the flask experiment where for heterotrophy, 16:2 does not appear until 48 hours. Under mixotrophy it is present from 12 hours (**Figure 3-9**). Both mixotrophy and heterotrophy in both media show an increase in 16:2 over time but its presence begins earlier in TAP mixotrophy than heterotrophy (**Figure 3-9**). In BBM medium mixotrophy gives higher levels of 18:3n3 at all timepoints, whereas in TAP, mixotrophy only gives higher levels of 18:3n3 in the first two timepoints only.

3.4.3 Different media

The two different media give dramatically different FAME compositions. There are differences in the shorter chain fatty-acids where BBMG has much more 14:1 than 14:0, and TAP has more 14:0 compared to 14:1. There is drastically increased 16:1 in BBMG than in TAP up to 24 hours. In BBMG there is significantly

more 18:2 than 18:3 but in TAP there is more 18:3 than 18:2. TAP always has more 18:3n3 than BBM.

3.4.4 The highest amount of 18:3n3 is found in TAP medium

ALA (C18:3n3) is the precursor to the omega-3 pathway and therefore it is desirable to have more levels of this as explained in the introduction. Levels of 18:3n3 seem to peak at 48 hours for BBM in both mixotrophy and heterotrophy, corresponding to a mol% of 26 % or 13 % respectively. In TAP the 18:3n3 level increases over time for heterotrophy settling at around 40 mol% at 48-120 hours hours,but decrease slightly in mixotrophy. This means that up to 12-24 hours, mixotrophic TAP cells contain more 18:3n3, however after 48 hours, heterotrophic TAP cells do. The maximum amount is 35-40 % 18:3n3 in TAP. There is more C18:3n3 in TAP than in BBMG.

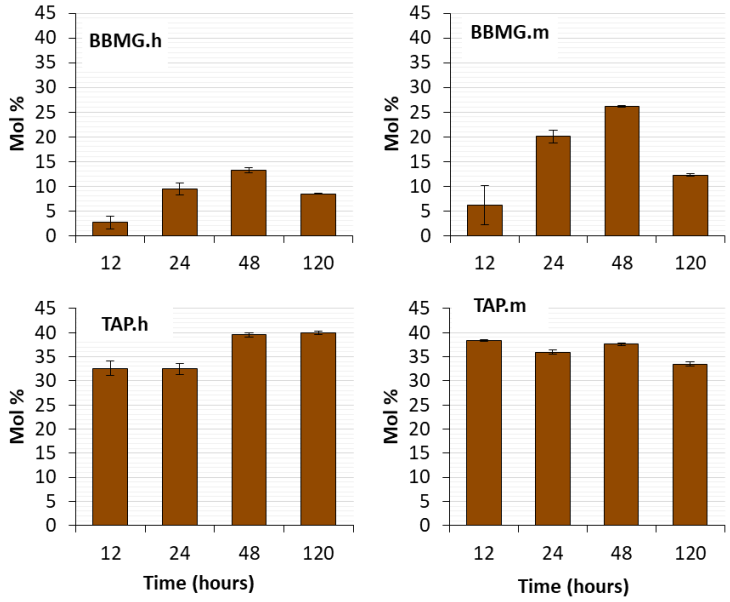


Figure 3-11 Changes of fatty-acid C18:3n3 over time in different media and growth modes. Standard error bars (n=3).

3.5 FAME and positioning profile of the TAG fraction of a 120 hour heterotrophically grown culture in BBMG.

Since TAGs are the preferred industrial class of lipids, it is interesting to examine the FAME profile of the TAG fraction (**Figure 3-7**). Also, stereospecific differences in the positioning of individual fatty acids on the glycerol backbone are interesting because they function differently in human metabolism (Karupaiah and Sundram, 2007; Zock et al., 1996) and in food applications certain positions are preferred, for example in baby milk formula C16:0 is preferred in the sn-2 position as this matches with the profile of human breast milk (Qi et al., 2018). Since one of the aims is to produce LC-PUFAs by engineering *C. sorokiniana* UTEX 1230, it is useful to explore this stereo-specificity. In this preliminary test, the TAG fraction from a 120 hour culture grown in BBMG heterotrophically was separately analysed. TAG scrapings from a TLC plate of a lipid extraction of *C. sorokiniana* had been prepared as described in the materials and methods and section 3.3.1. The TAG FAME and positioning analysis was done by Guillaume Menard at Rothamsted Research.

The TAG fraction is dominated by C16:0 and C18:2 fatty acids, and C18:3 is a smaller component with comparable levels to C18:1, C16:2 and C18:3 (**Figure 3-7**). A similar TAG FAME profile has been observed in previous work (Rosenberg et al., 2014). Large stereospecific differences are seen in fatty acids C16:0, C16:2, C16:3 and C18:2. With the exception of C16:0, the favoured position for the fatty-acid is in the sn-2 position. C18:3 occurs slightly more in the sn1 or sn3 position than the sn-2 position (**Figure 3-7**).

3.6 The effect of different carbon and nitrogen regimes on FAME profile, supernatant pH, neutral lipid content, biomass and cell density at a fixed timepoint

The manipulation of growth conditions influences the FAME profile and growth as seen in the previous section 3.4. In this experiment, differences are explored in more detail by specifically modifying the carbon and nitrogen sources in the same base media, because some interesting effects were observed in the

previous section which could have been attributed to this but needed further analysis in a more controlled experiment using a single base medium. The base medium chosen was TAP because this gave the better growth and C18:3n3 levels in the previous section. Due to the extent of the analysis done in this experiment, time constraints meant only a single timepoint was feasible. Measurements were recorded at 48 hours because out of the previous time-course experiments, it was one of the time-points which yielded the most C18:3n3. For the growth mode, mixotrophy was chosen because it gave a faster growth than heterotrophy.

3.6.1 Carbon and nitrogen source description

The base medium was the commonly used TAP medium which contains 20 mM Tris: 17.4 mM acetate as a buffering system. In conditions where the acetate is replaced by a different carbon source, the acetate is omitted from the base media and pH is instead adjusted with HCl. This does affect the buffering capacity of the media. Seven carbon sources (glycerol [Gly], glucose [Glc], acetate [Ac], xylose [Xy] and a combination of glucose with each of the other three sugars [GlyGlc, AcGlc, XyGlc]) and two nitrogen sources (ammonium [NH_4] and nitrate [NO_3]) were tested at two concentrations. These carbon sources were chosen because xylose and glycerol are common waste products of the paper and biofuel industries (Paranjape et al., 2016). Acetate and glucose are standard carbon sources and interesting effects have been observed with glucose before in terms of increasing TAG (Rosenberg et al., 2014).

For the carbon sources, the two concentrations tested are designated ‘normal’ and ‘high’ for the purposes of this experiment. The concentrations of each sugar were adjusted to account for differences in the number of carbon atoms (**Table 3-3**). The lower or ‘normal’ concentration is equivalent to 17.4 mM acetate, which is the normal concentration in TAP. The higher concentration was chosen to test the behaviour of the cells in a great excess of carbon, and was set to the equivalent of glucose at 500 mM (~90 g/L or 9 % (w/v)). This glucose concentration is nine times in excess compared to the carbon source used in BBMG in the previous section. When

comparing at a molar level, the excess between the 'normal' and the 'high' glucose is ~ 86 times atoms (**Table 3-3**).

For the nitrogen concentrations, 7.48 mM was used as the 'replete' condition since this is the concentration of ammonium in standard TAP. The 'low' nitrogen concentration is 0.748 mM, a ten-fold reduction which is representative of nitrogen starvation regimes commonly used (Ördög et al., 2016). The results from this experiment are divided in the following sections into observations on culture growth, culture pH, biomass and FAME profile.

Category	Condition name	Carbon source	Carbon conc. (mM)	Nitrogen source	Nitrogen conc. (mM)
"High" C	Ac	Acetate	1500	Ammonia (NH_4Cl)	7.48
	Gly	Glycerol	1000		
	Xy	Xylose	600		
	AcGlc	Acetate + Glucose	1500 + 500		
	GlyGlc	Glycerol + Glucose	1000 + 500		
	XyGlc	Xylose + Glucose	600 + 500		
	Glc	Glucose	500		
"Normal" C	Ac	Acetate	17.4		
	Gly	Glycerol	11.6		
	Xy	Xylose	6.96		
	AcGlc	Acetate + Glucose	17.4 + 5.8		
	GlyGlc	Glycerol + Glucose	11.6 + 5.8		
	XyGlc	Xylose + Glucose	6.96 + 5.8		
	Glc	Glucose	5.8		
"Normal" C, Low N	Glc(NH_4/10)	Glucose	5.8		0.748
"Normal" C, alt. N	Glc(NO_3)	Glucose	5.8	Nitrate ($NaNO_3$)	7.48
"Normal" C, Low alt. N	Glc(NO_3/10)	Glucose	5.8		0.748

Table 3-3 Selected carbon and nitrogen sources used in growth and lipid profile experiment on *C. sorokiniana* UTEX 1230. Concentrations of carbon sources were adjusted to account for the different number of carbon atoms in each compound so that that the were equivalent to 500 mM glucose in the 'High C' conditions or 17.4 mM acetate in the 'Normal C' condition. Abbreviations for the different conditions used throughout the text and figures are shown.

3.6.2 The ability to support cell growth varies greatly between carbon sources and concentrations

The final cell density was measured by haemocytometer at the harvest point of 48 hours and is shown in **Figure 3-12**, but it was beyond the scope of this experiment to monitor the growth curves of the cultures across the different conditions. The growth of the cultures is discussed below grouped into observations for high carbon conditions, low carbon conditions, and the nitrogen conditions.

3.6.2.1 At the 'high' concentrations, glucose is the only carbon source able to support cell growth

None of the cultures in the high sugar concentration actively grew from the initial 1 million cells/ml inoculation density apart from the culture supplemented with glucose ('High Glc') which rose to 13.5 × 10^6 cells/ml (**Figure 3-12**), just over half as much growth as for glucose at the 'normal' concentration. Cells in high glycerol ('High Gly') were sustained slightly above the inoculation density and cells in xylose ('High Xy') were sustained at the inoculation density. All other high carbon conditions displayed reduced density, probably from cell lysis (**Figure 3-12**).

3.6.2.2 At 'normal' concentrations, most carbon regimes are able to support cell growth but the level of growth varies depending on the carbon source

At the normal carbon concentrations, acetate + glucose ('Normal AcGlc') supported the highest cell density at 220 × 10^6 cells/ml. This was followed by normal concentration acetate ('Normal Ac') as the sole carbon source at 120 × 10^6 cells/ml (**Figure 3-12**). Normal AcGlc was 1.8 times higher than acetate on its own. The other normal concentration carbon sources were an order of magnitude lower at 11-31 × 10^6 cells/ml, apart from xylose ('Normal Xy') which did not support any cell growth (**Figure 3-12**). Although the cells do not increase in number from the initial inoculation density for normal xylose concentration, in the presence of xylose plus glucose the cells do grow ('Normal XyGlc'). In fact, the XyGlc culture grows to a marginally higher cell density than when glucose is the sole carbon source ('Normal'

Glc). 'Normal' concentration glycerol on its own grew the least, though this increases when combined with glucose.

3.6.2.3 Lower cell density is observed in the nitrate starved cultures but not ammonium starved cultures

For the nitrogen sources, which all had normal glucose concentration ('Normal Glc') as the only carbon source, replete nitrate ('NO_3') supported marginally more growth than replete ammonium ('NH_4') at 36 and 24 × 10^6 cells/ml respectively. However, at the low nitrogen ('$NH_4$1/10' or '$NO_3$1/10') concentration, cells grew slightly better on ammonium compared to nitrate at 24 and 12 × 10^6 cells/ml respectively (**Figure 3-12**). Interestingly, when ammonia is the nitrogen source, the final cell density is the same whether this ammonium is replete or at 1/10 concentration, both of which are lower than glucose with replete nitrate as the nitrogen source (**Figure 3-12**).

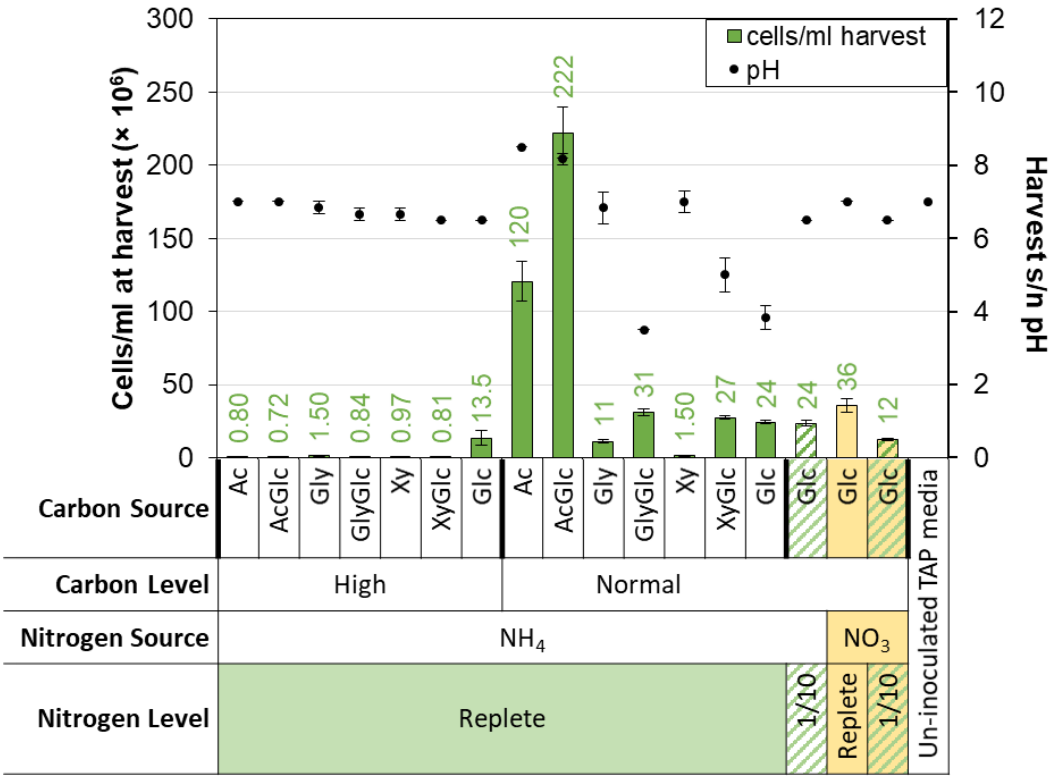


Figure 3-12 Cell density and culture supernatant pH of *C. sorokiniana* UTEX 1230 grown in different carbon and nitrogen conditions. Bars: cells/ml. Round markers: pH. Standard error bars (n=3). Carbon sources Acetate (Ac), Glycerol (Gly), Xylose (Xy), Glucose (Glc) and their combinations (AcGlc, GlyGlc, XyGlc) at concentrations equivalent by weighting to the number of carbons in the sugar to 500 mM Glucose (High) or 17.4 mM Acetate (Normal). Nitrogen sources are Ammonia (NH_4) or Nitrate (NO_3) at 7.48 mM (Replete) or stressed by reduction to 0.748 mM (1/10).

3.6.2.4 Pigmentation of culture varies depending on carbon source, even if final cell density is a similar value

The differences in cell densities can also be seen from the opacity and in some cases, colour, of the cultures (**Figure 3-13**). Though care must be taken with colour because changes in pigmentation can occur in actively growing cultures, especially when grown in heterotrophic conditions (Rosenberg et al., 2014). Cultures which are clear and are not opaque were ones which had low or reduced cell densities, for example high Ac, Gly, Xy, AcGlc, GlcGlc, XyGlc and normal Xy (**Figure 3-13**). However, the increased pale green colour in high Gly and slightly more opacity compared to the other non-growing high carbon conditions could be indicative of intact cells rather

than lysed cells, suggesting these cells are viable. The viability of these cells would have to be tested by trying to recover them into different growth conditions to validate this.

Thick, dark green cultures are clearly visible for the cultures with higher cell densities including high Glc and normal Glc (replete nitrogen only), Ac, GlyGlc, AcGlc and XyGlc (**Figure 3-13**). Interestingly, there is not much difference in final cell density between high concentration glucose and the low concentration glycerol (**Figure 3-12**), yet their colours are very different, with glycerol appearing paler green indicative of some bleaching (**Figure 3-13**).

Bleaching in actively growing cells is also seen when comparing replete vs 1/10 deprived nitrogen for both ammonia and nitrate (**Figure 3-13**), which is particularly striking considering for ammonia the cell density is the same for both the replete and 1/10 cultures (**Figure 3-12**), so the colour difference is not an artefact of cell density but of chlorophyll loss. Comparison of normal Gly and high Glc (replete ammonia) cultures also shows a marked colour change despite similar cell densities, with normal Gly appearing much paler, having a similar colour to the deprived nitrogen cultures.

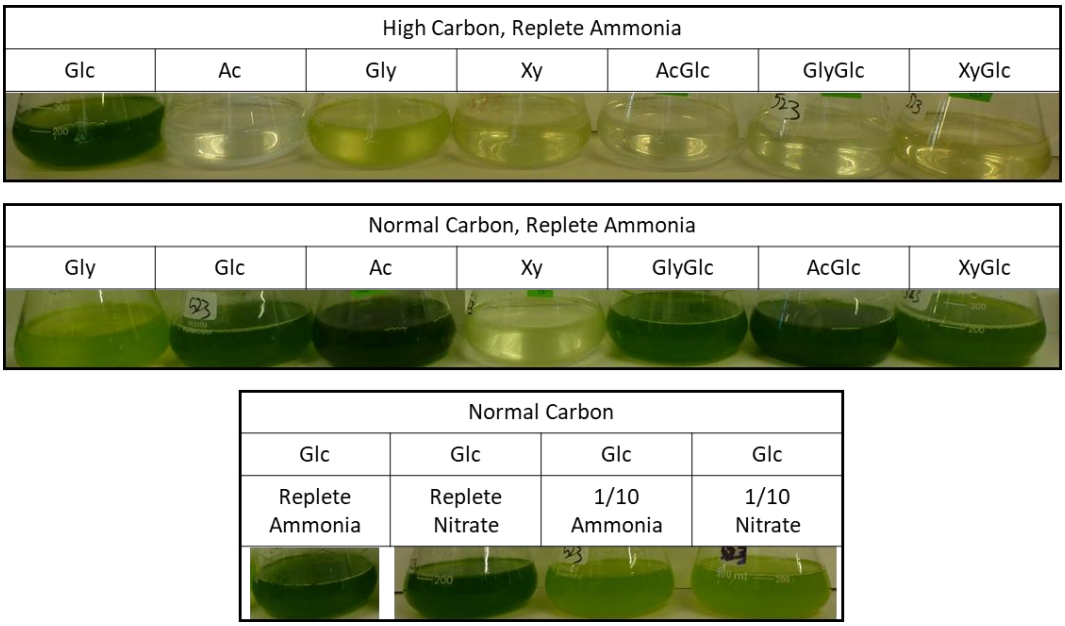


Figure 3-13 Colours of *C. sorokiniana* cultures grown in different carbon and nitrogen conditions for 48 hours. Cultures were grown in the same mixotrophic incubation conditions.

3.6.3 Culture pH is dramatically reduced in cultures containing glucose and higher in those with acetate

For those cultures that did not grow (all the 'high' carbon conditions except glc, and normal concentration xylose) the pH remained at 6.5-7.0 (**Figure 3-12**). A pH of 7 was also observed for normal glycerol, and the three normal glucose conditions with modified nitrogen regime ($GlcNO_3$, $GlcNO_3$1/10, $GlcNH_4$1/10).

In conditions containing glucose at normal carbon concentration with replete ammonia, the pH drops markedly from neutral to pH 3.5 (GlyGlc), 4, (Glc) or 5 (XyGlc), apart from when acetate is also present. The opposite occurs when acetate is one of the carbon sources: there is a substantial rise in pH for cultures containing acetate both on its own (Ac) and in the presence of glucose (AcGlc) from pH 7.0 to pH 8-8.5.

The drop in pH when there is no acetate is only seen in replete ammonia conditions. When glucose is the carbon source and the nitrogen source is nitrate, or 1/10 ammonia, then the pH remains 6.5-7.0. In the case of 1/10 ammonia compared to replete ammonia directly, the same cell density shows that the pH change is

independent of culture density and is related to the composition of the growth medium as discussed later. The impact of these pH changes could be having on the culture growth and subsequent lipid profile is discussed later.

3.6.4 The un-coupling of the correlation between freeze-dried biomass and cell density highlights differences in cell sizes depending on the growth regime

Freeze-dried biomass of the harvested culture was measured (**Figure 3-14**). Measurements of some samples were disregarded due to a confounding contribution of the high amounts of sugar in those samples relative to the final cell density.

The different carbon regimes resulted in different culture biomass at the harvest point of 48 hours. 'Normal' concentration acetate + glucose ('Normal AcGlc') produces the most biomass at over 200 mg, which is to be expected considering this condition had the most cells per mL (**Figure 3-12**). Most of the other cultures produced between 100 and 150 mg biomass (**Figure 3-14**). Cultures containing glycerol ('Gly') and xylose ('Xy') had very small total biomass at under 25 mg, consistent with their low cell numbers (**Figure 3-12**). These differences are expected due to the differences in final cell density (**Figure 3-12**). However, the magnitude of the differences cannot be fully explained by differences in cell density alone. For example, the acetate + glucose condition ('AcGlc') has almost 10 times more cells/ml than the glucose ('Glc') culture (**Figure 3-12**) yet not even twice of the biomass (**Figure 3-14**). As another example, acetate has 10 times more cells than Nitrate 1/10 glc (**Figure 3-12**) but only 1.5 times more biomass (**Figure 3-14**). Therefore, this suggests the cells are different sizes. Such differences in cell sizes were noticeable under the microscope during cell counting (data not shown).

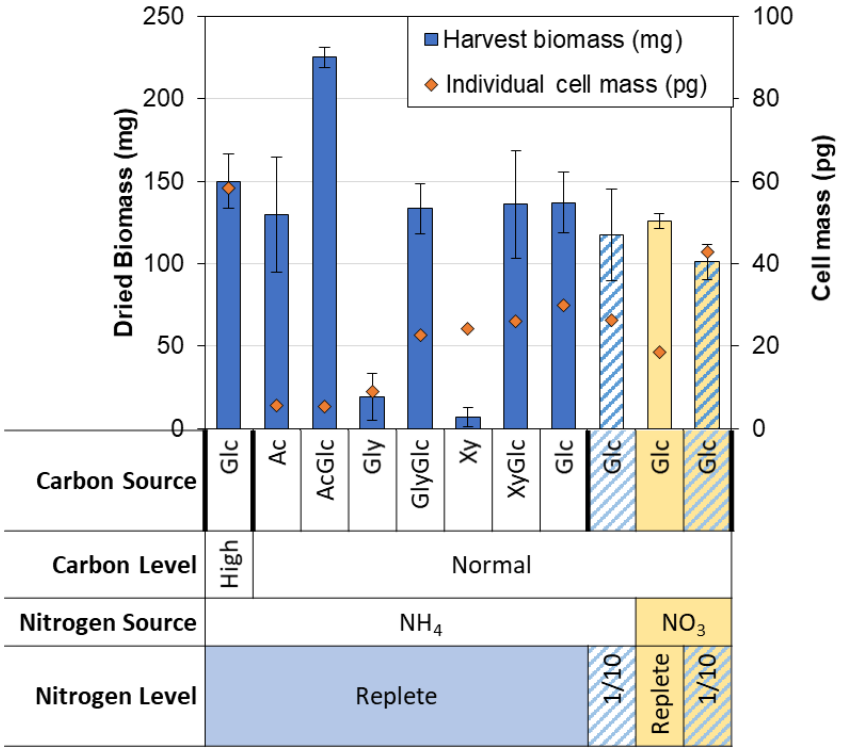


Figure 3-14 Total freeze-dried biomass and freeze-dried individual cell mass of *C. sorokiniana* after 48 hours growth on different carbon and nitrogen conditions. Standard error bars (n=3). Individual cell mass calculated from total culture volume (190 ml) and cells per ml at harvest.

As seen, the differences between the individual cell weights are not always reflected in the biomass. Looking at the individual cell mass helps to explain this. To calculate the individual freeze-dried cell mass (pg), the total culture freeze-dried biomass (mg) was divided by the total number of cells in the culture volume (190 ml) using the final cell density (cells/ml) (**Figure 3-14**). This highlights major differences between the different growth conditions and their effect on cell size. The acetate-only ('Ac') and the acetate + glucose conditions ('AcGlc') result in cells that are both under 5 pg each (**Figure 3-14**), whereas the glucose-only condition ('Glc') cells are 30 pg each which a 6-fold increase. Therefore, this approximately accounts for the 10 times more cells/ml of 'AcGlc' compared to 'Glc'. Therefore, the culture may contain 10-fold more cells, but the cells themselves are 86% smaller so the total biomass of the cultures appears more similar at only 1.5-fold increase.

Another example is that $GlcNO_3$1/10 has lower final cell density than replete GlcNO3 by three times (**Figure 3-12**), yet its total culture biomass is only marginally lower by a fifth (**Figure 3-14**). The difference in mass is seen in the mass of the individual cells themselves, with the 1/10 nitrate cells being twice as heavy as the replete nitrate (**Figure 3-14**) which could be attributed to TAG accumulation, as discussed later. This difference in cell weight in replete versus 1/10 nitrogen is not seen as much when ammonia is used as the nitrogen source (**Figure 3-14**).

3.6.5 The FAME profile is highly variable depending on growth regime and reveals insights into productivity of the different fatty acid species in response to different carbon and nitrogen sources

The cell densities, biomass and pH of the cultures was explored in the previous sections. This section onwards describes the lipids of the cultures. These are analysed by FAME (this section) for the total fatty-acid profile, and Nile Red assay in the next section for estimation of neutral lipid which can be an approximate indication of TAG levels.

The total FAME profile was measured from a sample of the culture taken at harvest (48 hours). Data was collected for cultures which grew or remained at the inoculation density: 'High' glucose (Glc) and glycerol (Gly), and all the 'Normal' carbon conditions (Ac, AcGlc, Gly, GlyGlc, Xy, XyGlc, $GlcNH_4$, $GlcNH_4$1/10, $GlcNO_3$, $GlcNO_3$1/10) (**Figure 3-15**). Owing to the inclusion of internal loading controls, the data was able to be expressed in a variety of ways. This was as fatty acid distribution per mol% of the cell (**Figure 3-15A**), fatty acid mass per cell (pg/cell) (**Figure 3-15B**) and fatty acid mass per culture (**Figure 3-15C**). The data for C18:3n3 was also separated for clarity as the focus of this analysis is to identify a condition which gives most C18:3n3 since this is the omega-3 pathway precursor. Depending on which way the data is expressed gives different conclusions (**Figure 3-17**) as discussed below.

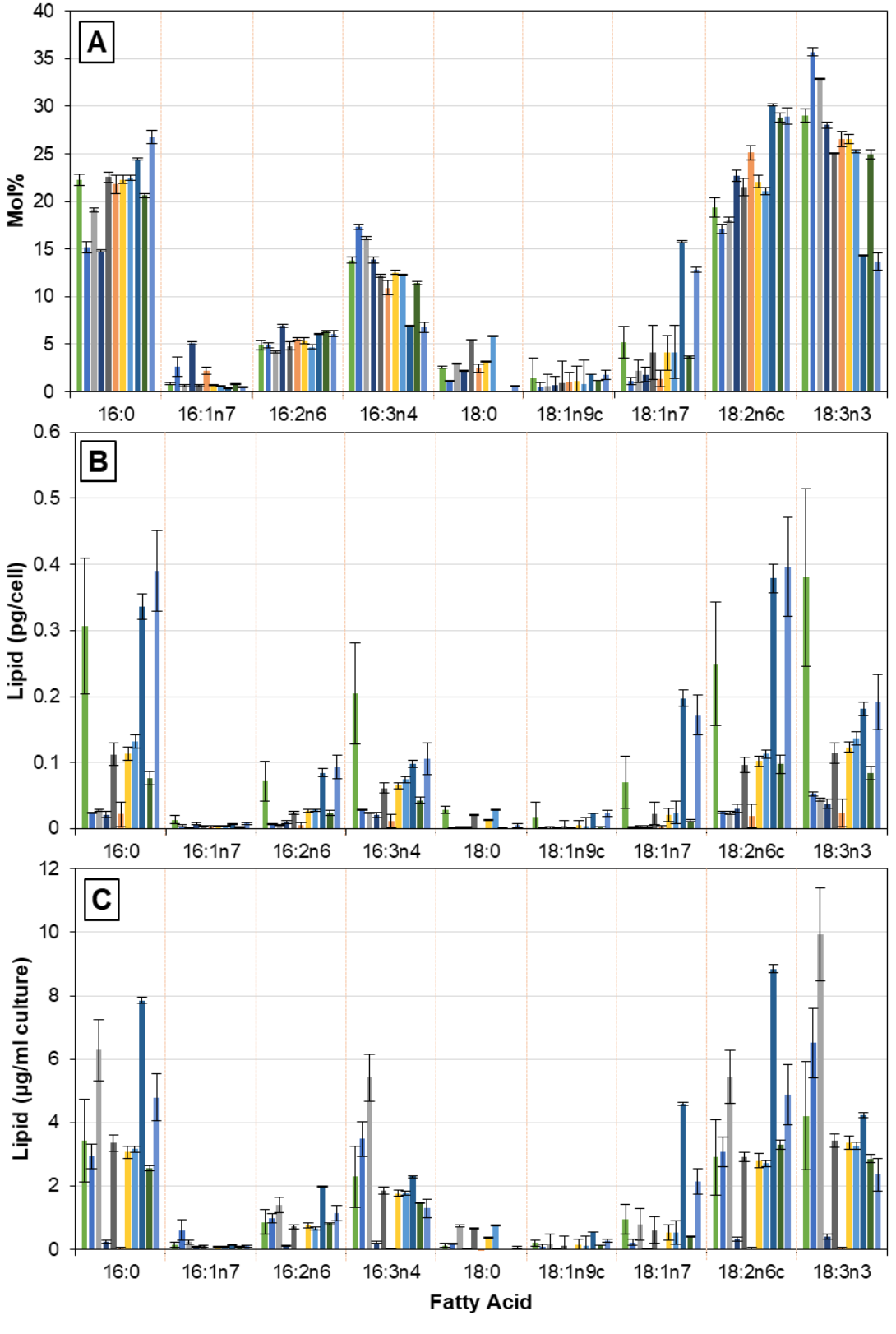

A
Mol%
40
35
30
25
20
15
10
5
0
16:0
16:1n7
16:2n6
16:3n4
18:0
18:1n9c
18:1n7
18:2n6c
18:3n3
B
Lipid (pg/cell)
0.6
0.5
0.4
0.3
0.2
0.1
0
C
Lipid (µg/ml culture)
12
10
8
6
4
2
0
Fatty Acid
Glc-High
Ac
AcGlc
Gly
GlyGlc
Xy
XyGlc
Glc NH4
Glc NH41/10
Glc NO3
Glc NO31/10

Figure 3-15 Total FAME analysis of wild-type *C. sorokiniana* UTEX 1230. Data expressed as mol% ratio (A), normalised to mass of lipid per cell (B) and total amount of lipid per culture (C). Data transformations of (B) and (C) used an internal C15:0 fatty acid standard. Fatty acids with very low amounts (e.g. under 2 mol% or equivalent) were excluded from the graph. SEM error bars (n=3) are shown.

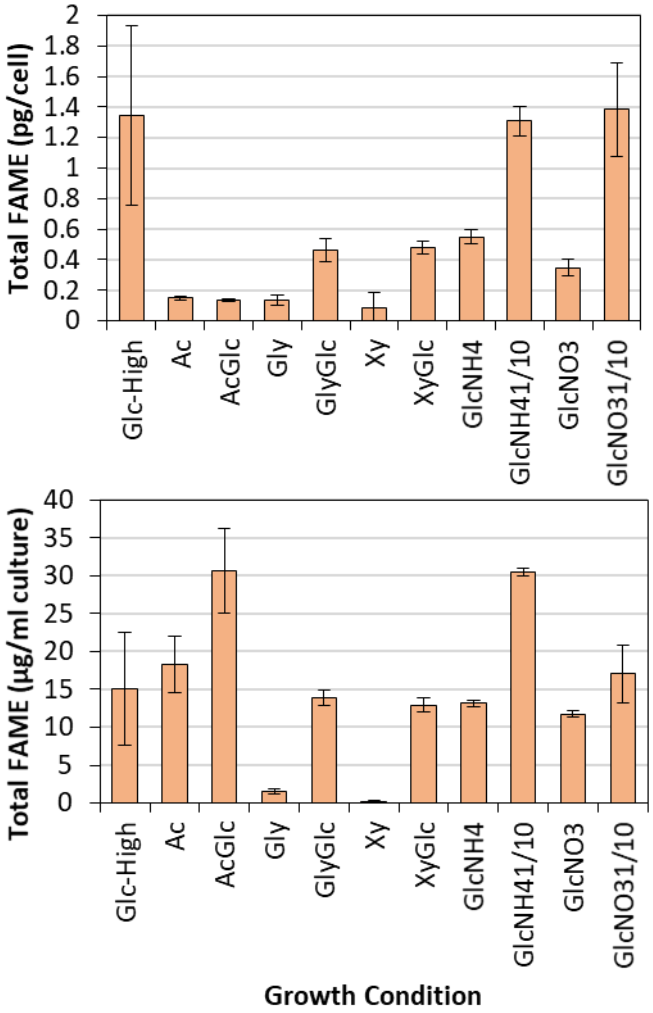


Figure 3-16 Total FAME across relevant acyl lengths of *C. sorokiniana* grown in different carbon and nitrogen sources. Data is normalised to cell number or culture volume.

Examination of the fatty acid profile using mol% reveals differences in fatty acid ratios between nutrient conditions (**Figure 3-15A**). In general, the profiles are dominated by 16:0, 18:2 and 18:3, but nitrogen starvation ('Glc($NH_4$1/10)' and 'Glc($NO_3$1/10)') causes a significant drop in the amount of C18:3n3 from ~25 mol% to 13%, and 16:3 from ~12% to 6%, but an increase in 18:1n7 from ~4 % to 12-16% and a small increase in 16:0 (**Figure 3-15A**). The nitrogen stressed cells also contain relatively more 18:2 than other conditions, reaching a maximum of ~30%. A notable perturbation is that replete nitrate condition ('Glc(NO_3)') also has this 18:2 increase whereas replete ammonium ('Glc(NH_4)') does not. The distribution of replete ammonium appears to contain more unsaturated fatty acids of 18:0 and 16:0 than replete nitrate instead of increased 18:2 (**Figure 3-15A**).

If the lipid mass per cell (pg/cell) is considered (**Figure 3-15B**) then for each individual media condition the ratio between the fatty acids is consistent with the mol% data, but it allows for interesting differences in lipid content to be seen between the growth conditions. Growth conditions of high glucose ('Glc-High'), and nitrogen starvation ('Glc($NH_4$1/10)' and 'Glc($NO_3$1/10)') have drastically more lipid per cell than the other growth conditions (**Figure 3-16**). This is particularly clear in the fatty acid species which dominate the profile during these growth conditions, notably C16:0, 16:3, 18:2 and 18:3 for high glucose ('Glc-High', although the error bars are quite substantial) and 16:0, 18:1 and 18:2 for nitrogen starved. It appears cells grown in lower nitrogen sources preferentially synthesize C16:0 and C18:2 lipids. Also, when observed under a microscope, the cells grown on high carbon ('Glc-High') were notably larger than in other growth conditions (data not shown). Growth conditions with notably low amounts of lipid per cell include carbon source of acetate, acetate+glucose, glycerol and xylose (**Figure 3-15B** 'Ac', 'AcGlc', 'Gly' and 'Xy').

Looking at the data in terms of ng lipid per ml culture allows the productivity and growth rate of the culture to be considered because the fatty acid profile is normalised to the harvest cell number of the whole culture after the 48 hours cultivation time (**Figure 3-15C**). For example, despite both the nitrogen starved cultures having similar mass of 18:2, 18:1 and 16:3 and 16:0 per cell (**Figure 3-15B**), the whole culture lipid yield for these fatty acids is lower in the starved nitrate

condition ($GlcNO_3$1/10) than the starved ammonia condition ($GlcNH_4$1/10) (**Figure 3-15C**). This is related to the differences in final cell densities and final biomass from these conditions: the final cell density for starved nitrate was lower than starved ammonium (**Figure 3-12**) and although total biomass was similar, the nitrate starved had a larger individual cell mass than ammonium starved (**Figure 3-14**). For certain fatty acid species such as 16:0 and 18:0, the lipid productivity of the nitrogen starved cultures impressively rivals that of the best biomass producing culture 'AcGlc normal concentration' (**Figure 3-15C**).

3.6.5.1 Closer examination of the C18:3n3 FAME profile compared across the different cultures

The data for the LC-PUFA omega-3 precursor C18:3n3 (ALA) is isolated for clarity in **Figure 3-17** sinc it is the level of this fatty acid which is desired to be maximised for strain engineering**.** If the ratio of fatty acids is considered via mol% then acetate at 'normal' concentration ('Ac') has the highest proportion of C18:3n3 at 36 mol%, followed by acetate + glucose ('AcGlc') at 33 mol% (**Figure 3-15A, Figure 3-17**). Looking at lipid pg/cell, glucose at high concentration ('Glc-High') has the largest amount of C18:3n3 per cell at 0.37 pg/cell (though the error bar is quite substantial), twice as much as the next best conditions of low nitrogen ('Glc($NH_4$1/10)' and 'Glc($NO_3$1/10)') at 0.18-0.19 pg/cell (**Figure 3-15B, Figure 3-17**).

Despite the overall large amount of total lipid per cell for High-Glc and low nitrogen (**Figure 3-16**), it can be seen that AcGlc and Ac have the highest amount of C18:3n3 per culture at ~10 µg/ml and ~6.5 µg/ml, respectively (**Figure 3-17**). This is compared to 'Glc-high' and the low nitrogen conditions yielding between 2.5-4.5 µg/ml (**Figure 3-15C**). It seems the higher cell density for AcGlc compared to Ac and other conditions is enough to counteract the effect of a higher mol% C18:3n3 in Ac (**Figure 3-15A, Figure 3-17**), and the higher lipid mass per cell of the Glc-High and low nitrogen conditions (**Figure 3-15B, Figure 3-17**). Though, the two low nitrogen conditions actually show high amounts of C18:3n3 considering the reduced mol% level, but they preferentially accumulate other fatty acids 16:0, 18:1 and 18:2 instead (**Figure 3-15**).

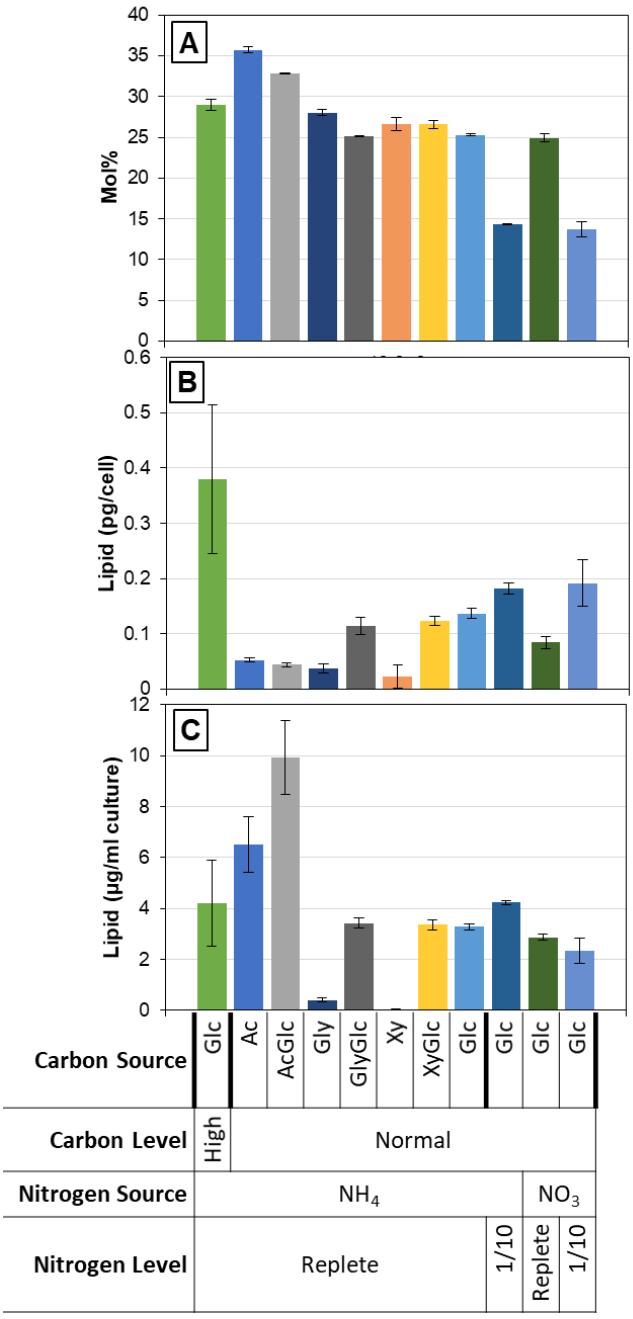


Figure 3-17 Changes in amount of C18:3 (ALA) from the FAME profile of *C. sorokiniana* UTEX 1230 grown in different carbon and nitrogen regimes. The data is expressed as mol % (A), pg/cell (B) or ng/ml (C) culture volume.

3.6.6 Neutral lipid appears to accumulate to high levels in many growth regimes and appears to be dependent on multiple different factors

In addition to exploring the total fatty-acid composition of the cells, it is of interest to explore the effect these growth conditions have on the amount of TAG in the cells because this is the desired form of lipids for industry. Nile Red is a lipophilic fluorescent dye which binds intracellular neutral lipids and other hydrophobic compounds, resulting in a shift in emission wavelength when bound (Greenspan and Fowler, 1985). It can therefore be used as an approximate indicator of the amount of TAG in algal cells, since this is a neutral lipid. The raw Nile Red data is a complex emission spectrum and these are discussed first in section 3.6.6.1. Subsequently, using the extracted data for neutral lipid, patterns and correlations with growth conditions are discussed in section 3.6.6.2.

3.6.6.1 Examination of the raw curves

For samples of live *C. sorokiniana* culture supplemented with 10 % DMSO, binding of Nile Red to lipids results in a shift of the broad emission peak from 653 nm to 586 nm (Vonlanthen, 2013). In this study Nile red fluorescence was measured using excitation at 510 nm with scanning emission at 550-740 nm. The resulting profile is a curve consisting of multiple peaks which can be seen in the representative samples shown in **Figure 3-18** and are summarised in **Table 3-4**. The raw curves (**Figure 3-18** top row) were normalised to cell number to give the fluorescence per cell (**Figure 3-18** bottom row), and the fluorescence at 586 nm was extracted as the relative level of neutral lipid within the cell for comparison across the growth regimes (**Figure 3-19**).

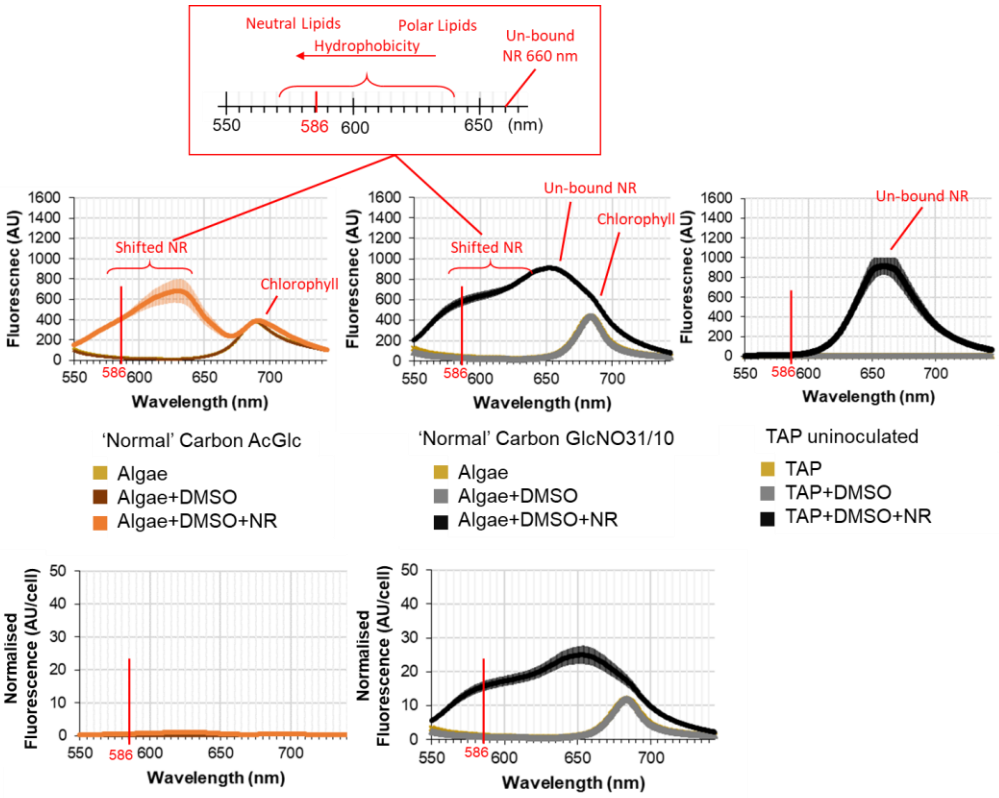


Figure 3-18 Example of emission spectra for *C. sorokiniana* cultures stained with Nile Red lipophilic probe. Representative curves are shown for two of the growth conditions and the uninoculated TAP+DMSO control. The excitation wavelength was 510 nm. The multiple peaks are explained within the text. The emission wavelength of 586 nm which was used as the representative wavelength for neutral lipid is highlighted on the scale. The raw curves (top row) were normalised to the number of cells in the cuvette (bottom row) for the final data.

Peak	Description
685 nm	Chlorophyll autofluorescence
660 nm	Nile red in plain TAP media with DMSO
580-650	Blue-shifted Nile Red peak: smaller wavelengths correspond to increasingly hydrophobic environments (i.e. neutral lipid compared to polar lipid)
586 nm	The wavelength chosen to represent neutral lipid, in accordance with previous work (Vonlanthen, 2013).

Table 3-4 Description of the peaks occurring in the emission spectrum of *C. sorokiniana* cells stained with Nile Red lipophilic dye. Corresponding representative curves can be seen in **Figure 3-18**.

In the fluorescence spectrum of cells stained with Nile Red and excited at 510 nm, there are multiple peaks but one peak is a well-defined, quite broad peak at 685 nm which corresponds to chlorophyll auto-fluorescence (**Figure 3-18**). This peak was can also be seen in cells + DMSO controls in section 3.3 **Figure 3-8**. The other peaks are all related to Nile Red fluorescence. At 660 nm there is a very broad peak which is un-bound Nile Red in TAP medium plus DMSO (**Figure 3-18**). When cells are present in the sample, this Nile Red curve shifts to smaller wavelengths (is blue-shifted) because of the presence of lipids and other hydrophobic compounds. For cells which do not contain much neutral lipid, this shift seems to typically manifest as a very broad peak at ~630 to 645 nm (**Figure 3-18**). The neutral lipid peak is broad and occurs around 585-590 nm, but rather than a defined peak it often appears as a shoulder to the neighbouring peak (**Figure 3-18**).

In order to analyse the data, the signal from the cuvette was normalised to signal per cell by dividing by the known number of cells in the cuvette. This helps to mitigate the effect of any background non-bound Nile Red on the signal. The normalised fluorescence signal at the neutral lipid peak of 586 nm is shown in **Figure 3-19**. Although this Nile Red analysis was performed for all growth regimes, it is clear that the data is not meaningful for those samples which did not grow, as the only peak is the un-bound Nile Red. Therefore, these data were excluded from the final analysis.

3.6.6.2 Comparison of the neutral lipid signal per cell between growth regimes highlights associations between neutral lipid accumulation and low pH, high sugar or nitrogen deprivation

The data from 586 nm for the non-excluded conditions, which should correspond to neutral lipid (and hence is an approximate indication of TAG), was isolated for comparison in **Figure 3-19**.

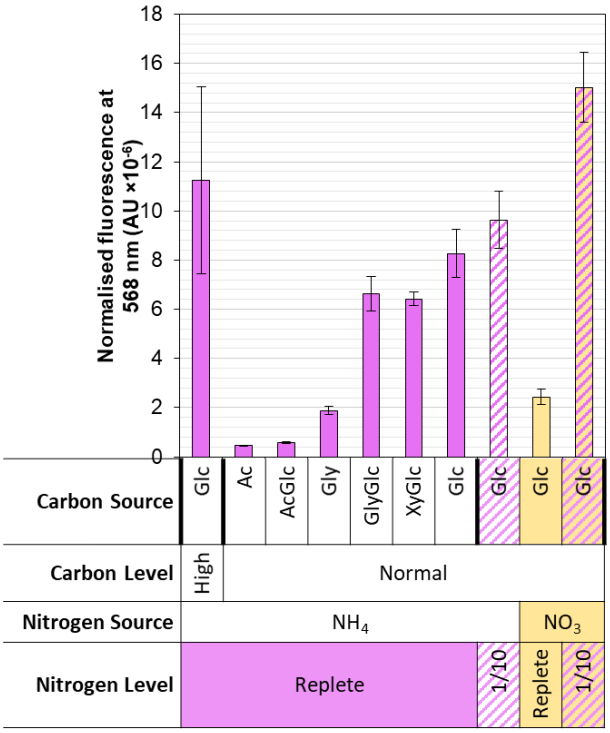


Figure 3-19 Nile Red fluorescence at 586 nm of *C. sorokiniana* grown in different carbon and nitrogen regimes. Standard error bars (n=3). The signal was normalised to the number of cells in the cuvette so is indicative of the neutral lipid content per cell.

Growth regimes which gave the two highest levels of neutral lipid are 'High' Glc, and 'Normal' Glc with 1/10 nitrate (Glc$NO_3$1/10). The next highest amounts of Neutral lipid occur in GlyGlc, XyGlc, GlcNH_4 and Glc$NH_4$1/10. Those with the least neutral lipid are Ac and AcGlc followed by Gly then GlcNO_3 replete.

High levels of neutral lipid are expected in the two 1/10 nitrogen conditions because stressing cells in this way is a commonly used method to induce TAG in microalgae. More surprising is the relatively high amount of neutral lipid in many of the other non-nitrogen stressed growth regimes, namely GlyGlc, XyGlc and GlcNH_4 replete, which have neutral lipid levels comparable to Glc$NH_4$1/10 (**Figure 3-19**). This suggests there must be some other factor causing stress to the cells and hence inducing neutral lipid accumulation. One notable correlation is that these cultures have a low pH of 3.5 to 5.0 compared to the others which are maintained at near neutral, or above (max. pH 8.5) (**Figure 3-12**). It seems that this low pH is a possible stressor to the algal cells causing a similar neutral lipid phenotype as seen in cells

under nitrogen stress. This is particularly exemplified by highlighting the accumulation of Nile Red signal in both the ammonium replete ($GlcNH_4$) and ammonium deprived ($GlcNH_41/10$) cultures to unexpectedly similar levels, whereas for replete or starved nitrate ($GlcNO_3$ or $GlcNO_31/10$) only the nitrate starved culture accumulates Nile red signal as would be expected. Therefore, the difference in lipid accumulation between the replete nitrate and replete ammonium culture could be related to the lower culture pH value of ~ pH 4 for ammonium compared to pH 7 for nitrate. This suggests low pH could be able to induce neutral lipid to levels comparable to that of nitrogen depletion.

A final condition with increased neutral lipid is the high glucose ('High Glc') condition (**Figure 3-19**). In this case, there is no correlating pH change as with the other examples; instead the pH of this culture was at pH 6.5 (**Figure 3-12**). This culture was not under nitrogen stress either. Therefore, the stress of the osmotic pressure from the high sugar concentration (500 mM or 9% w/v) may be the cause of neutral lipid accumulation. Due to the inability of other high carbon conditions to sustain growth, the data in this experiment does not include another example of this. A worthwhile future experiment would be to test different concentrations of other carbon sources or a combination of them to increase the osmotic pressure, to an extent which does not kill the cells. It is also important to consider that since the Nile Red data is expressed per cell, then cell size can affect the reading because larger cells with have more physical space to accumulate neutral lipids.

3.7 Discussion

This chapter is concerned with investigating factors affecting growth and lipid accumulation in *C. sorokiniana.* The main aim was determining an appropriate growth condition which preferentially accumulates the omega-3 precursor fatty-acid C18:3n3. After examining the effect of temperature on *C. sorokiniana* growth, a multitude of growth conditions were also examined by FAME profile including time-course, trophic mode, base media and finally a screen of several nitrogen and carbon sources at two different concentrations. In the latter experiment, a preliminary assay of neutral lipid content via Nile Red fluorescence was carried out because this

information is industrially relevant due to neutral lipids being preferred for extraction and processing in some applications.

The results of the experiments highlight the complexity of the cellular lipid response to different environmental factors and the importance of considering culture productivity alongside the lipid profile. C18:3n3 is preferentially accumulated to levels of 33 mol% or 10 µg/ml culture when acetate and glucose are provided as carbon sources. Analysis of the Nile Red data identified acidic pH stress as a promising avenue for further research into neutral lipid accumulation microalgae. This is applicable to all industry fields interested in maximising lipid production, from biofuels to the production of specific high-value omega-3 oils.

3.7.1 *Chlorella sorokiniana* growth and cell mass in multiple growth regimes

In agreement with literature reports (Li et al., 2013; Sorokin, 1959; Sorokin and Krauss, 1959), *C. sorokiniana* exhibits shorter doubling times at increased temperatures. In this experiment, the doubling times were ~3.5 hrs at 35 °C and ~7 hrs at 25 °C (**Figure 3-2**). The temperature of 25 °C was chosen as it is representative of the standard temperature conditions used in our lab, and 35 °C represents the thermophilic condition. However, 35 °C is not the upper temperature limit of growth, with studies reporting growth up to 42 °C (Li et al., 2013; Sorokin, 1959; Sorokin and Krauss, 1959). The ability of *C. sorokiniana* to reach stationary phase in under 24 hours, at least in these growth conditions which may not be optimal (for example, no CO_2 gassing of the cultures was carried out), is advantageous for industry where high growth rates are valued, especially in continuous culture. One drawback of this experiment though is that because it used the Algem photobioreactor caution must be taken when labelling the growth phases because the automated OD reader of the Algem becomes saturated quickly and in order to compare these growth curves to those obtained by other means they must be calibrated.

3.7.1.1 *C. sorokiniana* is able to grow at high growth rates under mixotrophy and heterotrophy

When comparing mixotrophy and heterotrophy, mixotrophy supported higher growth but this could be because the heterotrophic cultures ran out of carbon. This would have to be confirmed by running additional experiments on the spent media to characterise its components, for example by HPLC. As seen in one study (Rosenberg et al., 2014), heterotrophy can support high cell densities overall which would be beneficial for the production of bioproducts, especially if this could be maintained in continuous mode of operation.

3.7.1.2 Carbon and nitrogen regimes impact culture growth

The inability of any of the high carbon conditions to support cell growth, except high glucose, suggests that these are toxic to the cells at this concentration and that glucose is the only carbon source out of these which is suitable for supplementing cell cultures at high concentration. A further experiment could be carried out to identify maximum concentration for each of the carbon sources, since some of them are beneficial feedstocks since they are waste carbon sources: for example, glycerol which is a by-product of biodiesel production (Yang et al., 2012). Examination of growth in true heterotrophic conditions (i.e. growth in complete darkness) rather than mixotrophy (carbon source and some light) would also be a valuable test because heterotrophy has also been reported to support higher cell densities over longer growth periods (Rosenberg et al., 2014).

For conditions where cell growth was supported, analysis of their growth curves would be needed to determine whether differences in harvest cell densities were due to varying growth rates (i.e. cells had not reached maximal stationary phase cell density by 48 h), or if some carbon sources were facilitating higher maximal cell densities by reducing the need for photosynthesis and hence any disadvantages of high density cell-shading. Additionally, the effect of the growth condition of seed inoculum must be considered: seeds were grown in standard TAP medium, so there may be a period of adaptation to the new media which has not yet been characterised. The differing pigmentation of *Chlorella* sp. cultures has been observed

previously in heterotrophy (Rosenberg et al., 2014), under high light (Grudzinski et al., 2016) and nitrogen deprivation (Vonlanthen, 2013).

3.7.1.3 The effect of carbon source and concentration on cell mass

It makes sense that the total biomass from the cultures is different considering their different cell densities. However, the difference in biomass is not directly correlated with cell number. This incongruousness is due to variance in individual cell mass as seen in **Figure 3-14**. One important consideration is that the cells in different growth conditions were probably all in different growth stages as no full growth curve was monitored and only a single harvest time point was measured due to time constraints. Additionally, the cultures were not synchronous. Considering cell size and morphology may change throughout a growth cycle (Chioccioli et al., 2014), this could explain some of the variation. The cell size may be relevant to the extraction of the desired product because it is costly to process biomass and break open the tough cell wall of *Chlorella* sp. Larger cells may be more amenable to cell disruption but it would be worth consulting with appropriate industrial research groups to determine how much of an economic factor this may be. The increase in cell size of *C. sorokiniana* when grown on glucose has been observed in previous work (Shihira and Krauss, 1965) and also the related microalga *Chlorella vulgaris* has cell size and shape dependant on the culture conditions (Chioccioli et al., 2014).

3.7.1.4 Xylose may be an interesting carbon source to pursue in further research

Xylose is relevant sugar because it is an important part of hardwood hemicellulose and there is lack of pentose sugar utilisation for biofuel (Leite et al., 2016). It has previously been used to induce lipid accumulation in microalgae (Leite et al., 2016). In the experiments in this chapter, xylose had an interesting observation. When 'normal' concentration xylose was the sole carbon source, cells appeared unable to divide, being sustained at just above the inoculation density, but in the presence of glucose they grew again, and this growth was more than glucose on its own (**Figure 3-12**). The ability of the cells to grow when xylose was combined with glucose suggests that xylose is not toxic to the cells at this concentration. However, the lack of growth beyond the inoculation density suggests that xylose is unable to

be utilised by the cells and that additionally for some reason cells were unable to utilise the light and grow via photosynthesis. It would be interesting to run a parallel experiment in autotrophic conditions to validate this hypothesis.

Previous studies (Hassall, 1958) have found that D-xylose arrests cell division at a concentration of 0.5% in flasks bubbled with CO_2, but not those with an alternate carbon source such as glucose, fructose or mannose, with or without CO_2. This corroborates the findings presented in this chapter where *C. sorokiniana* is unable to grow on xylose only, but when a second carbon source is added, such as glucose, they are able to grow. The same previous study (Hassall, 1958) observed that the addition of glucose to bleaching cells in xylose-containing medium after 4 days allowed the cells to regenerate and multiply in 48 hours, suggesting death is caused by starvation rather than a toxicity mechanism. They propose that xylose or a derivate is blocking a part of the Calvin cycle and thus photosynthesis. If xylose is a true blocker of photosynthesis, it could be a useful tool in research.

However, contrary to the observations in this chapter and the previous report, one study shows that the UTEX 1602 strain of *C. sorokiniana* can uptake D-xylose, but that in the presence of D-glucose its uptake is inhibited (Zheng et al., 2014). The differences here highlight the difficulties in comparisons in microalgal work where there are such large variations between strains of the same species, which is especially true in *Chlorella sp.* where many unknown algae of great diversity were classified as such (as discussed in introduction section 1.5.1).

3.7.1.5 Different base media affects cell growth

There were interesting variations in cell growth in BBM and TAP base media. The main constitutional difference between these two media are their carbon and nitrogen sources, however they also differ in other elements (**Table 3-5**) which could have contributed to the growth variation. For example, there is half as much potassium in TAP compared to BBM but slightly more phosphates in BBM than TAP. The sulphates are at similar levels but the iron is in a different ionic form. Also, TAP media uses trace element solution optimised for the green alga *Chlamydomonas*

reinhardtii which was shown to increase algal growth (Kropat et al., 2011) whereas BBM does not (Nichols and Bold, 1965).

The effect of these media differences could be tested by conducting an experiment with TAP containing the alternate iron source, or a more general test using the different trace element solution. Modifications of phosphates, iron and sulphates levels have been used to induce lipid accumulation in microalgae previously (El-Sheekh et al., 2015; Otsuka, 1961) so it would be interesting to investigate the effect of this on lipid accumulation in *Chlorella sp.* as well. Calcium, magnesium and zinc have been shown to largely contribute to differences in growth rate in a large screen of media (Wolf et al., 2015). Since the presence of different ions ("ionic context") and micronutrients has been shown to affect growth and lipid content, high-throughput optimisation approaches would be a valuable way to properly determine optimal growth media rather than using commonly used media for convenience (Evens and Niedz, 2010; Hallenbeck et al., 2015a, 2015b; Wolf et al., 2015). However, the cost of making a tailored medium would need to be balanced with the change in output of the desired product.

Component	BBM	TAP
Major Ions (mM)		
$(Ca)^{2+}$	0.170	0.340
EDTA	0.171	0.058
$(Mg)^{2+}$	0.304	0.406
K^{+}	2.70	1.602
Acetate	-	17.420
Tris	-	20.0
$(SO_4)^{2-}$	0.36	0.408
$(Cl)^{-}$	0.78	8.234
$(PO_4)^{3-}$	1.72	1.000
Na^{+}	3.37	0.160
$(NO_3)^{-}$	2.94	-
$(NH_4)^{+}$	-	7.478
$(CO_3)^{2-}$	-	0.022
Trace Ions (uM)		
$(Mo)^{6+}$ or $(MoO_4)^{2-}$	4.93	0.200
$(SeO_3)^{2-}$	0.0116	0.10
$(Zn)^{2+}$	30.67	2.50
$(Mn)^{2+}$	7.28	6.00
$(Fe)^{2+}$	17.91	-
$(Fe)^{3+}$	-	20.00
$(Cu)^{2+}$	6.29	2.00
$(Co)^{2+}$	1.68	-
$(Sn)^{4+}$	0.0038	-
$(VO)^{2+}$	0.0101	-
I^{-}	0.02	-
$(BO_3)^{3-}$	184.700	-
$(Ni)^{2+}$	0.013	-

Table 3-5 Molar composition of the two different growth media used in this work. Molar composition was calculated from recipes given for the various media. The BBM medium is that of Sigma (B5282) and the TAP medium is standard but uses the new 'Special K' trace nutrient solution, not Hutners like traditional TAP.

3.7.2 Suitable conditions for preferential C18:3n3 accumulation

C. sorokiniana has a variable lipid profile and some conditions favour the production of PUFA precursors. In the course of the work presented in this chapter, several potential parameters were investigated in terms of their effect on *C. sorokiniana* growth and fatty-acid profile (**Figure 3-1**). Exploration of mixotrophy and heterotrophy in two different media determined that mixotrophic growth in TAP was preferential for C18:3n3 accumulation in terms of proportion of all the fatty-acids across the cell and that C18:3n3 level was relatively consistent over time but peaked

at 48 hours during early stationary phase (**Figure 3-11**). The growth condition that produced the most C18:3n3 in the preliminary work was heterotrophy in TAP at 48 or 120 h harvest which gave almost 40 mol% 18:3n3. This was closely followed by mixotrophy in TAP at 12 or 48 h harvest which gave 36-38 mol% 18:3n3 (**Figure 3-9**).

This second growth condition (mixotrophic TAP) was chosen for the more in-depth study into different carbon and nitrogen sources presented in this report. 48 hours cultivation was chosen because at this timepoint cultures accumulated more 18:3n3 in the initial experiments and is practical to work with in the lab. In the data presented, the maximum mol% of C18:3n3 occured is the standard TAP ('normal Ac') condition at 36 mol% C18:3n3 (**Figure 3-17**). In addition to exploring the proportion of different FAMEs using mol%, other measures utilising quantification by an internal GC standard and normalisation to individual cells or total culture were presented. The maximal condition in each of these data sets is different: for mol%, the top conditions are Ac (36%) followed by AcGlc (33%); for pg lipid per cell it is 'high' Glc followed by both low 1/10 nitrogen conditions (0.38, 0.17-0.19 pg/cell respectively) and for lipid per culture it is AcGlc at 9.8 μg/ml culture and Ac at 6.5 μg/ml culture (**Figure 3-17**). The mol% value is in agreement with the value of 36-38% obtained in the preliminary study for these conditions (**Figure 3-9**), suggesting the methodology is reliable in this case. The optimal condition was a balance between optimal fatty-acid profile ratio, total lipid accumulation, and cell density/growth rate. Although the amount of C18:3n3 per cell was low for culture conditions of combined AcGlc, the increased biomass of these cultures at the harvest timepoint means that they are the most productive producers of C18:3n3 per culture volume at 9.8 μg/ml culture.

The consideration as to whether it is more beneficial to have a higher ratio of C18:3n3 per cell or a higher total yield per culture may also depend on the enzyme kinetics. C18:3n3 is the precursor of the omega-3 pathway on which transformed heterologous enzymes will act upon (see introduction **Figure 1-10**). It follows logically that for maximum product (omega-3 LC-PUFA) accumulation, the proportion (i.e. mol%) of the C18:3n3 substrate would need to be maximised in the cell. However, it might still be better to consider maximal total mass of C18:3n3 per cell rather than relative proportion, even if the proportion of C18:3n3 is lower compared to other

fatty acids, because the enzyme:substrate ratio may be more optimal in this dynamic. This would depend on the specificity and K_M of the enzyme: if the enzyme is promiscuous or has a low K_M then the growth condition with the highest mol% would be preferred, since only a small amount of substrate is needed for the enzyme to reach its V_{max}, and it would be important that the enzyme was more likely to react with the desired substrate (C18:3n3) than an unspecific substrate. If the enzyme is highly specific or has a high K_M, maximal C18:3n3 mass per cell may be preferred as it maximises the C18:3n3 substrate per unit enzyme in the algal cell. These different options highlight the need to consider enzyme kinetics such as the k_{cat}/K_M ratio on an individual basis when deciding what level of substrate to aim for in the cell, and if these kinetics are uncharacterised for the enzymes in question, then multiple growth conditions may need to be considered.

3.7.3 Other interesting FAME patterns

Other differences and patterns in FAME patterns were recorded in section 3.4 and 3.6.5. Changes in lipid profile are expected over time as the cell remodels itself throughout its growth cycle. In order to explain the changes in abundance of various fatty acid chains, one can explore the function of the different fatty acids. Though it is worth noting that the cultures are not synchronised in this experiment and in order to examine the lipid transfer between classes during the growth cycle then synchronous culture would have to be used. This is one benefit of using microalgae as models to explore these physiological processes, because they can be easily synchronised by light/dark regimes (Sorokin and Krauss, 1959; Tischner and Lorenzen, 1979).

In section 3.5 is an example of the fatty acid profile of just the TAG fraction of UTEX 1230 cells and their breakdown into sn position on the TAG molecule (**Figure 3-7**). Different positioning can indicate whether the fatty acids in the TAG molecule have been generated by the prokaryotic or eukaryotic pathway and so enable insights into the cellular metabolism. This kind of thorough analysis would be a beneficial research avenue in future characterisation of the growth conditions identified as high C18:3n3 accumulators in this work. A previous study has analysed the FAME profile

of TAG in UTEX 1230 grown in BBM supplemented with glucose at 10 g/l and does have some similarities with the results presented here. For example, they observe almost ten-fold more FAME 18:2 compared to 18:3 in heterotrophy and ~5 times more in mixotrophy. In their FAME profile, 16:0 and 18:2 have similar amounts to each other. Over time, 16:0 and 18:2 show the most increases whereas 18:3 stays almost the same for both mixotrophy and heterotrophy as measured from three to thirteen days (Rosenberg et al., 2014).

3.7.4 pH and osmotic stress as alternate triggers for inducing neutral lipid accumulation in *Chlorella sorokiniana*

Cell density or growth rate may also have been affected by the pH changes seen in the cultures. Cultures containing glucose underwent acidification, except those that also contained acetate or had high/low nitrate or low ammonium in the media (**Figure 3-12**). The precise impact these pH changes have on cell growth would require further monitoring of the growth curves because in this experiment the changes in final cell density reflect the influence of both the nutrient conditions, growth phase and pH interaction. However, the identical final cell density of replete and 1/10 ammonia for Glc normal suggests it is attenuated to some degree. pH crash of algal cultures is a known issue in industry, and cells do eventually kill themselves when grown on ammonium which is why it is sometimes called ammonium 'toxicity' although the ammonium itself is not toxic (Scherholz and Curtis, 2013). Extra buffering against pH changes, for example via bubbling of carbon dioxide, could also aid higher cell densities. However, this adds to the cost and complexity of bioreactor design (Tan and Adebusuyi, 2016). Therefore, approaches that carefully tailor the growth medium to consider proton stoichiometry, sometimes with multiple nitrogen sources, have received recent attention (Scherholz and Curtis, 2013; Wang and Curtis, 2016).

A common way to induce TAG accumulation in microalgae is through the use of nitrogen deprivation (de Lomana et al., 2015; Fan et al., 2014; Zhang et al., 2017) but the lower productivity of the algal cells means it is not always a viable strategy. There have been two-stage growth regimes developed to mitigate this but these have

additional draw-backs such as needing to change media which is expensive and impractical especially at large scale (Minhas et al., 2016). Although the purpose of the carbon and nitrogen screen was mainly focussed on identifying suitable conditions for C18:3n3 accumulation, the examination of the neutral lipid data led to some interesting insights for avenues to be investigated further in terms of triggers for neutral lipid accumulation in *C. sorokiniana*.

The results in this chapter show promising preliminary data that acidic conditions (pH 3.0-4.5) or osmotic stress by high glucose (500 mM or 9% w/v) can also induce neutral lipids. These stresses could be industrially relevant because they are cheap, simple and reproducible and can be triggered by the preferable process of addition of a factor to the growth media, rather than removal of a component which can result in costly processing to change media.

3.7.4.1 Why the pH varied in the cultures

Acidification of algae cultures is typically attributed to the H^+ ion release associated with the ammonium uptake symporter (Scherholz and Curtis, 2013; Von Wirén et al., 2000). The minimised acidification in the nitrate or low ammonium growth conditions in this experiment supports this explanation. Nitrate uptake is typically associated with a rise in the pH of algal culture (Scherholz and Curtis, 2013; Wang and Curtis, 2016). For acetate-containing conditions, acetate is being consumed by the algae which typically pushes the Tris:acetate buffer system to more alkaline because the level of the acidic component is reduced. It is interesting that the assimilation of acetate in these media mitigates the ammonia-induced acidification enough to cause a pH increase. This may suggest that either acetate is the preferred carbon source in dual carbon source conditions, and in fact the second carbon source is not being utilised. Or, that the presence of acetate allows utilisation of the second carbon source (e.g. glucose) to a greater extent by mitigating the deleterious drop in pH. To test this, monitoring the amount of each carbon source left in the system would be necessary, either by assay or some other analytical method such as HPLC.

In conditions where acetate was not one of the carbon sources, the media was brought to correct pH with HCl instead. Since HCl is a strong acid rather than a weak acid like acetate, it reduces the buffering capacity the medium. The use of a different base medium such as BBM (used in the growth conditions test in section 3.2.2.2, **Figure 3-4**) which does not use the carbon source (acetate) as part of a buffering system, would allow the effect of these dual carbon sources to be analysed without the need to consider a secondary effect on pH.

Furthermore, it is interesting to note that in *Chlorella sp.* the rate of glucose uptake has been shown to be pH dependent: the hexose transporter uptakes a proton and when the proton concentration in the media is reduced then levels of carbon uptake also diminish (Komor and Tanner, 1974). Perhaps the addition of glucose could be used as a strategy to avoid pH increase on ammonium growth on unbuffered media, if acidic pH is not desired. It would be interesting to determine whether the absence of a pH decrease in the High-Glc condition is because the cells were not growing enough to metabolise the ammonia to cause a pH change, or whether the large amount of glucose present meant the uptake of protons mitigated this effect. To test this, the carbon uptake from the media would need to be measured and for the lower carbon conditions, investigating whether there is a timepoint at which the glucose has been totally depleted which then corresponds with a subsequent pH decrease.

3.7.4.2 Preliminary evidence for the use of acidity as a neutral lipid trigger in *Chlorella*: correlations between high Nile Red signals and acidic pH of the cultures

In the Nile Red (NR) data from the carbon and nitrogen regime experiment in section 3.6.6, initially it was surprising that so many of the non nitrogen-deprived cultures had a high NR signal. For example, apart from the nitrogen starved $GlcNH_4$1/10 and $GlcNO_3$1/10, high NR signal was also recorded in 'Normal' Glc, XyGlc, GlyGlc and 'high' Glc. Since algal cells typically accumulate lipid under stress, it follows that the presence of neutral lipid can itself be an indicator of cell stress and so identifying the cause of the stress by closer examination of the other measured

parameters revealed some interesting correlations. 'Normal' concentration Glc, GlyGlc and XyGlc all had acidification of media reaching pH 3-3.5 for Glc and GlyGlc or 4.5 for XyGlc. The case is particularly evident when directly comparing the Nile Red signals of the replete and starved nitrogen conditions for ammonia ($GlcNH_4$ and $GlcNH_41/10$) and nitrate ($GlcNO_3$ and $GlcNO_31/10$). As expected, both the 1/10 (deprived) growth regimes have high neutral lipid regardless of whether the nitrogen source is nitrate or ammonia. However, the only replete nitrogen condition with high neutral lipid is $GlcNH_4$ which has a pH of ~4 compared to pH 7 for replete $GlcNO_3$. This suggests that pH is able to stimulate neutral lipid accumulation to a similar extent as nitrogen stress.

3.7.4.3 Osmotic stress as a trigger for neutral lipid accumulation

Although High Glc did not cause acidification of the media, the cells grown in this condition were larger so would have been physically able to contain more neutral lipid per cell. Also, the high 9% w/v carbon could have placed the cells under osmotic stress. The addition of glucose supplementation causing lipid accumulation is in agreement with literature reports (Rosenberg et al., 2014). However, glucose is a relatively expensive sugar compared to waste carbon sources such as xylose and glycerol, so it could be worth trying to focus efforts on optimising growth on these carbon sources instead. Though, this would be more relevant for the production of low value products like biomass feed or biofuel, since if the desired product is of high-value, such as an omega-3 PUFA, then there is more tolerance in the cost of supplied nutrients to be economically competitive.

3.7.4.4 Further experiments

In order to confirm the effect of pH and osmotic stress, additional controlled experiments would need to be carried out. In the literature, alkalinity has been shown to induce TAG accumulation in *Chlorella sp.* (Guckert and Cooksey, 1990; Rai et al., 2015). In terms of osmotic pressure, 1-2% glucose has been reported as the optimum osmotic concentration (Hassall, 1958), whereas in this experiment 9% was used. The data in this chapter has provided some interesting preliminary observations. It would be beneficial to repeat this work and conduct it using a Design Of Experiment

methodology, which would be more rigorous and has been successfully used to explore the parameter space for conditions for heterotrophic cultivation of algae (Chiranjeevi and Venkata Mohan, 2016; Wolf et al., 2015). The 'normal' carbon Ac and AcGlc conditions clearly are not stressed as reflected in their high cell density and low neutral lipid signal. It would be interesting to take these optimal growth conditions and stress the cells using addition of acid or high carbon to see if neutral lipid accumulates.

A limitation in this experiment is that only one time-point was sampled and so differences in growth phase were not taken into account: the cells were likely growing at different rates, but the growth curves were not measured because it was beyond the scope of this preliminary experiment. Therefore, there may not have been enough time for lipid droplets to accumulate to the maximum level, since *Chlorella sp.* cultures are often grown for 5 days or more (well into stationary phase) before assessing lipid content (Rosenberg et al., 2014; Vonlanthen, 2013), and nitrogen deprived cells typically start accumulating lipids at the late log phase. Characterisation of lipid at different timepoints would be required to identify the growth condition which provided an optimal balance between cell stress causing TAG accumulation, and any deleterious effect on growth rate/biomass production.

3.7.5 Other explorations of the neutral lipid data

3.7.5.1 Insights into lipid remodelling from neutral lipid

In addition to quantification of total lipid per cell from FAME analysis with internal standards, neutral lipid per cell was also measured via the Nile Red (NR) assay, which is commonly used as a proxy for TAG accumulation. Although the main focus of this work was to identify a condition with maximal C18:3n3, TAG accumulation was also explored because enrichment of TAGs is desirable for industrial extraction of lipids. When the NR data is examined in combination with the FAME lipid pg/cell data, it is possible to gain insights into any lipid remodelling which may be occurring in the cell, and whether this is directed to TAG or other areas such as phospholipid membranes.

In some cases, all of the change in FAME lipid per cell can be attributed to accumulation in TAG. For example, there is a ~2-fold increase in FAME lipid per cell in nitrate deprived ($GlcNO_3$1/10) culture compared to the nitrogen replete culture ($GlcNO_3$) but a ~5-fold increase in NR signal for neutral lipid per cell. In other conditions the change in FAME lipid per cell is not reflected in the NR signal. For example, cultures grown in high concentration glucose (high $GlcNH_4$) condition had a ~2-fold increase in FAME lipid per cell but only a ~1.1 fold increase in NR signal compared to normal glucose ($GlcNH_4$) (**Figure 3-15**, **Figure 3-19**). These discrepancies could be due to variations in lipid type remodelling between the different growth conditions and could be tested by examination of the different lipid classes by GC-MS or TLC.

3.7.5.2 Nitrogen stressed cells FAME profile

Examining the Nile Red data in combination with the lipid profile gives some preliminary insight into avenues and targets which could be pursued in future experiments. As expected, the nitrogen stressed cells do appear to contain more neutral lipid than other growth conditions, but they also preferentially accumulate different fatty acids. The nitrogen stressed cells accumulate unsaturated fatty acids which are good for a biofuel application, but not for the omega-3 precursor 18:3n3. Even though nitrogen starved cells had higher lipid content per cell, the higher culture productivity and preferential accumulation of 18:3n3 in Ac and AcGlc growth conditions compensates for this and results in accumulation of more 18:3n3 omega 3 precursor overall in the culture (at least for the timepoint of 48 hours used in this experiment). Therefore, for production of PUFAs the use of nitrogen stress as a lipid induction technique may be counter-productive, even if total lipid content of the cells is higher. This chapter highlights that a better strategy to explore in further experiments would be the use of exogenous carbon supplementation such as acetate and glucose because although individual lipid content per cell is not higher than nitrogen stressed cells, there is a preferential accumulation of PUFA and greater productivity. This hypothesis would need to be validated in the context of the whole growth curve of the cells, because as shown in section 3.4.1, the FAME profile changes over time but the carbon and nitrogen regime data presented in section 3.6

represents only a snapshot of the cells at 48 hours in an undefined growth phase for most media as the growth curves were not monitored.

3.7.6 Methodology issues in algal growth and lipid studies

Recently, there has been increasing attention in both the media and within the scientific community that there is a problem with replicability of research. One survey found that when scientists attempted to reproduce another researchers experiment, 70% failed, with 50% even failing when trying to reproduce their own work (Baker and Penny, 2016). Algal research is no stranger to this either. Analysis of the composition of algal biomass is arguably one of the most important aspects of algal laboratory work considering that a large proportion of algal research is focused on their potential as feedstocks for biofuels or nutritional components. However, in published research there is great variation between the methods used and their effectiveness varies between the algal species used (Cavonius et al., 2014; Dejoye Tanzi et al., 2013; Laurens et al., 2012; Yan Li et al., 2014; Tang et al., 2016b). For example, in lipid analysis many publications use derivatives of traditional methods such as Bligh and Dyer or Folch-based solvent extraction, but do not fully specify their modifications leading to difficulties with replicability and comparison of work between different laboratories (Iverson et al., 2001). The adoption of technical standardisations is the norm in industry and would be worthwhile to facilitate inter-laboratory studies (Laurens et al., 2017).

The potential sources of error in lipid analysis work can be overwhelming and there is often a lack of detail in methods sections of published articles. Considering in some research, the inability to replicate findings was distilled down to differences in the way researchers stirred their cells (Hines et al., 2014), the omission of even a centrifuge speed or an incubation time in a protocol may have repercussions on an experimental result. In this chapter, several challenges in algal measurements were revealed and are some potential sources of error as discussed below.

3.7.6.1 Measurements of cell density and biomass

In this chapter, cell growth was measured by OD_{750} and manual cell counting for cell density. OD_{750} is a measure of biomass rather than cell number because it is affected by the size of the particle. Although it is possible to make calibration curves between cell density and OD_{750} (Ding et al., 2016) they are unsuitable if the cell changes significantly in morphology or size. Discrepancies between OD_{750} and cell density were observed in **Figure 3-3** where the increase in manual OD_{750} is more than cell number and **Figure 3-2**. OD_{750} in the Algem bioreactor has its own issues because there are physical limitations to the automated system where the sensor gets saturated because the path length is across the base of the flask, and therefore a calibration curve must be done using manual measurements. However, these can be inaccurate if there are changes in cell morphology or size. Hence for the experiments in different media and carbon and nitrogen sources, cell density was used.

Additionally, the use of optical density as a way of sharing data between labs is not ideal because due to variations between spectrophotometers and plate readers it is not an absolute measurement like counted cell density. These problems have been highlighted in bacterial work, especially in relation to cell fluorescence (Beal et al., 2016). However, counting cells with a haemocytometer, although an absolute value rather than an indirect measurement, the method is subject to large error, especially considering the equipment was designed for counting large red blood cells which are much larger than *C. sorokiniana*, and is incredibly time consuming. An alternative to this would be to use automated machines, though there are additional costs. Other possibilities are the use of software and algorithms which can analyse images and perform the counts which would be the ideal solution (Geissmann, 2013; Grishagin, 2015). Other alternatives also include using colour based analyses to estimate biomass (Sarrafzadeh et al., 2014), but this would be inappropriate in this case because of the bleaching observed in some growth conditions despite similar cell densities (**Figure 3-13**).

3.7.6.2 Total Lipid extractions

In this chapter, two solvent based methods with different mechanical disruption and phase separation were tested. Mechanical disruption has been shown to be important in algal lipid extraction (Kim et al., 2016; Rakesh et al., 2015) but in this case there were no obvious differences in yield between the two methods (**Table 3-2**).

There are many lipid extraction techniques with a multitude of minor, often unspecified modifications with little standardisation. Partly this is because a lot of the techniques need calibrating to individual organisms themselves. Although this is adequate when doing in depth studies on just one strain, it is less useful for studies assaying differences between many strains such as bioprospecting or exploring culture collections.

3.7.6.3 FAME analysis for acyl chain analysis

The analysis of chain length by derivatising the sample for Fatty-Acid Methyl Esters (FAME) is a common way to analyse algal lipids but there are some drawbacks such as high variability between experiments. Additionally, separation of the FAME mixture by Gas Chromatography with flame ionisation detection (GC-FID) is labour and time intensive and requires calibration to identify the peaks based on standard FAME mixes, some of which have only become available within the past few years. An alternative to identifying the peaks is by GC-MS (Gas chromatography-mass spectroscopy) but again this requires access to such equipment. However, FAME analysis can be a powerful quantitative analysis when standards of known concentration are used.

In FAME analysis it is common practise to remove uneven fatty-acids 15:0 and 17:0 from the profile because they are regarded as contamination from bacteria (Lang et al., 2011; Viso and Marty, 1993). Therefore, in axenic cultures these species should not be present, and this is why C15:0 was chosen as a standard. However, if there are large amounts of these bacterial signals then it raises issues as to whether

the bacteria are contributing to the FAME profile as well and therefore providing unreliable or confounding variable.

3.7.6.4 Nile Red as an estimate of neutral lipid

Nile Red is a commonly used method for estimating neutral lipid in microalgae (Alemán-Nava et al., 2016; Chen et al., 2009; Halim and Webley, 2015; Morschett et al., 2016). However, the technique is sensitive to multiple variables which must be optimised for each strain (Alemán-Nava et al., 2016; Pick and Rachutin-Zalogin, 2012; Rumin et al., 2015). Suitable conditions for *C. sorokiniana* had been previously identified (Vonlanthen, 2013) but the resulting scanning emission curves are complex and do not show clearly separated peaks due to the presence of polar and non-polar lipids (**Figure 3-18**).

The response of Nile Red to both neutral and polar lipids was described during the first report of the compound where Nile Red was described as a "hydrophobic probe in which the fluorescence maxima exhibit a blue-shift proportional to the hydrophobicity of the environment" (Greenspan and Fowler, 1985). Although Nile Red preferentially binds to neutral lipid, it will bind to polar lipids in excess and these different peaks have been widely reported (Halim and Webley, 2015; Pick and Rachutin-Zalogin, 2012; Rumin et al., 2015).

In this work, a single wavelength was extracted from the emission curves to represent the neutral lipid peak. Although this is the accepted way of doing it, examination of the curves highlights the inaccuracy because this peak is often a shoulder to the remainder of the un-bound Nile Red or the polar lipid peak from excess Nile Red. Therefore, in future it could be worth adding less Nile Red, but this creates other problems such as quicker Nile Red saturation. Alternatively, the data from the entire emission spectrum could be used, and perhaps it would be more representative to use a measure of the extent of the peak shift to indicate the amount of neutral lipid. Some studies have used more complicated mathematical analysis to deconvolute multiple curves (Viseu et al., 2003).

3.8 Conclusion

Lipid metabolism is part of central cell metabolism as therefore is a complex process affected subtly by multiple input parameters, as overviewed in the introduction. In line with the aims of the thesis to engineer *C. sorokiniana* to make high value oils such as the omega-3 fatty acids EPA (20:5n3) and DHA (22:6n3), this chapter aimed to probe a specific part of *C. sorokiniana* lipid metabolism: the accumulation of the fatty acid ALA (18:3n3) in the lipid profile. ALA is the longest fatty acid chain which *C. sorokiniana* synthesises and is the precursor for any longer omega-3 fatty acids which the cell could be genetically engineered to synthesise.

This chapter found that the lipid profile of the cell exhibits changes in response to trophic mode, base media and cultivation time. The results presented show that a suitable growth condition for maximising 18:3n3 is a balance between achieving a desireable FAME profile at an individual cell level, and the overall productivity of the cultures. Among the conditions screened, the most suitable include: heterotrophic or mixotrophic growth in unmodified TAP media which yielded 40 or 36-38 mol% 18:3n3 (36 mol% corresponded to 6.5 µg 18:3n3/ml of culture) respectivley at harvest points of 48 and 120 hours; and TP media with acetate (17.4 mM) and glucose (5.8 mM) which yielded 33 mol% or 9.8 ug 18:3n3/ml of culture. Although nitrogen stressed cells appeared to contain more total lipid, the FAME profile of these cells tended towards increased unsaturated fatty acids which are more suited to biofuel applications, and shows that the use of nitrogen stress as a trigger for PUFA accumulation could be counter-productive. For TP media with glucose at 500 mM, the largest amount of 18:3n3 per cell was obtained (0.38 pg/cell) but cells had lower productivity by the harvest timepoint of 48 hours (4.2 µg 18:3n3/ml of culture). However, investigation into this condition with a longer cultivation time could be valuable because the growth phase of the culture was not known and the FAME profile did not show any unfavourable shifts to saturated fatty acids (18:3n3 was 29 mol%).

Having identified some suitable growth conditions for PUFA production, the subsequent chapters in this thesis are concerned with the genetic engineering of the

microalga *C. sorokiniana*, which is required in order to introduce heterologous elongase and desaturase enzymes to act upon 18:3n3 fatty acids and produce longer omega-3 fatty acids. The growth conditions determined in this chapter could be used to maximise PUFA production in this genetically engineered strain.

CHAPTER 4 OPTIMISATION OF A TRANSFORMATION STRATEGY FOR *CHLORELLA SOROKINIANA*

4.1 Introduction

Although *C. sorokiniana* UTEX1230 is a promising strain in terms of its fast growth rate, lipid profile and TAG accumulation, the currently limited ability to introduce heterologous enzymes to allow production of specific bioproducts is important in its potential as an industrial strain. Genetic engineering of microalgal strains for heterologous gene expression or up/down expression of native genes is seen as one way of achieving production of specific chemicals as an alternative to screening for natural high producers, and allows some non-natural products to be produced. Specifically of interest in this project are genes to modify the lipid pathways to allow high value PUFAs to be made. This genetic engineering strategy requires a way of transforming microalgal cells and selecting for those containing recombinant genes. Although a preliminary investigation into using electroporation in *C. sorokiniana* was carried out (described in section 4.2), this chapter focuses on using *Agrobacterium*-mediated transformation, chosen as there was a precedent for this method working from previous studies (Hiegle, 2014). Based on a review of published protocols of *Agrobacterium*-mediated transformation across a range of microalgae, and examination of the infection mechanism, rational modifications to the initial protocol are suggested. The antibiotic sensitivity of wild-type *C. sorokiniana* is also rigorously assessed.

4.1.1 The use of G418 and zeocin as selection agents and the test plasmid used in initial transformations

Antibiotic selection is one of the most common ways of detecting transformants as this involves dominant positive selectable markers and although varying levels of expression may be observed due to positional effects, selection ultimately gives a binary output: cells live if they contain the construct and die if they

do not. There are many antibiotic or herbicide resistance genes that have been used in microalgae such as *AphVIII* for paramomycin and neomycin, *Hpt* for Hygromycin B, *NptII* for G418 (Geneticin), *Ble* for Bleomycin/Zeocin™ and *Als* for suphonylurea herbicides (Reviewed in: (Bashir et al., 2016; Gangl et al., 2015; León-Bañares et al., 2004; Radakovits et al., 2010)). Some, such as *aadA* for spectinomycin and streptinomycin resistance are specific to the chloroplast (Bashir et al., 2016; Gangl et al., 2015). Different species of microalgae vary greatly in their natural resistance to these, so they must be tested for each strain.

C. sorokiniana UTEX 1230 is a notably tough strain and the antibiotics Hygromycin B and Paramomycin, which are commonly used in other microalgae, are unsuitable for selection in *C. sorokiniana* due to growth of the wild-type at unreasonably high concentrations (Vonlanthen, 2013). G418 and Zeocin™ have previously been found to be effective (Hiegle, 2014; Vonlanthen, 2013) and the RbleC23 plasmid construct containing both G418 and zeocin resistance cassettes has been successfully utilised for *Agrobacterium*-mediated transformation of *C. sorokiniana* (Hiegle, 2014). Plasmid RbleC23 (**Figure 4-1**) uses the pCAMBIA2300 backbone (GenBank AF234315.1) suitable for *Agrobacterium*-mediated transformation of plant cells. Between the right border and the left border (which define the region of DNA that *Agrobacterium* transfers to its host) are two antibiotic resistance cassettes oriented in diverging directions. The first, the *nptII* gene for resistance to G418, is used as a selectable marker for isolation of putative transformants, and the second, the *ble2I* gene which confers resistance to zeocin, is used as a reporter cassette to score and confirm transformants. The *nptII* CDS is driven by a duplicated CaMV 35S promoter and terminator (CaMV35S2) and the *ble2I* CDS is driven by the *C. reinhardtii RBCS2* promoter and terminator. *Ble2I* is a modified version of the *ble* gene, containing two endogenous *C. reinhardtii* introns from the *RBCS2* gene, which has been shown to increase protein expression in *C. reinhardtii* (Lumbreras et al., 1998).

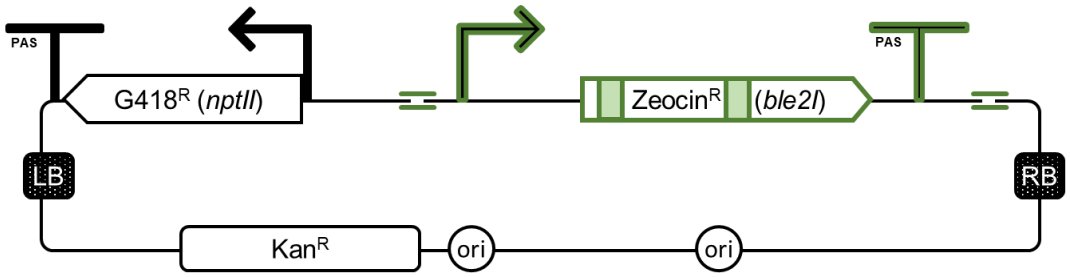


Figure 4-1 Diagram of plasmid RbleC23 used in initial *Agrobacterium*-mediated test transformations of *C. sorokiniana*. The plasmid is a total of 10kb. It uses the pCAMBIA2300 vector backbone (GenBank AF234315.1) modified by cloning a Zeocin resistance cassette (Hiegle, 2014). The construct between the left border (LB) and right border (RB) sequences is 3779 bp and consists of a G418 resistance gene *nptII* and Zeocin resistance gene *ble2I*. These genes are flanked by the duplicated Cauliflower Mosaic Virus 35S promoter (CaMV25S2) and single terminator sequence/poly-A signal (35ST) or the *C. reinhardtii* RuBisCo promoter and terminator (*RBCS2*) respectively. The Zeocin resistance gene is modified to include two copies of *C. reinhardtii RBCS2* intron1 (green bands). The backbone has two origins of replication, one for *E. coli* and one for *Agrobacterium*.

G418 was first reported for use in transformation of *Chlorella* strain ATCC-22521 by Hawkins *et al*. (Hawkins and Nakamura, 1999) where they used 1mg/ml. The encoding gene, *nptII* has since been reported as used for other members of the *Chlorella* genus (Liu and Chen, 2016; Yang et al., 2016). In *Chlorella,* the use of *ble* for bleomycin or Zeocin™ resistance has only been reported for *Chlorella ellipsoidea* and in one study they used 10 µg/ml for selection (Liu and Chen, 2016; L. Liu et al., 2013; Yang et al., 2016). However, due to large phenotypic differences between species in the same genus, this level is not necessarily transferable to *C. sorokiniana*. Indeed, previous studies used Zeocin concentrations of up to 350 µg/ml (Hiegle, 2014). The use of the bacterial *ble* marker gene is well established in *C. reinhardtii* which is the first microalga for which it was used (Lumbreras et al., 1998; Stevens et al., 1996).

4.1.2 Transformation methods in microalgae

Transformation methods used for nuclear transformation of algae include microparticle bombardment (biolistics), electroporation, agitation with glass beads or silicon carbide whiskers, aminoclay nanoparticles, *Agrobacterium*-mediated

approaches, polyethylene glycol (PEG) mediated DNA uptake, and episome transfer by bacterial conjugation (Gangl et al., 2015; Karas et al., 2015; Liu and Chen, 2016). Some of these techniques such as PEG-mediated methods often require the generation of protoplasts by removal of the cell wall, and although not a strict requirement, in some cases this may increase the efficiency of other methods such as electroporation and agitation with glass beads. Each method has different advantages or disadvantages in terms of specialist equipment required, cost, time and efficiency of transformation (Gangl et al., 2015).

Most nuclear transformations in microalgae result in random insertion of the DNA into the genome since illegitimate recombination processes dominates over homologous recombination pathways present in the nucleus; in fact efficient homologous recombination of nuclear transgenes has only been reported in two algae species, both of which are red algae (Doron et al., 2016). Although this is actually a useful feature for applications such as creating random mutant libraries, for genetic engineering it is a disadvantage because positional effects hinder precision by impacting transgene expression and disrupting host genes (Hlavova et al., 2015). However, it may be possible to increase the number of homologous recombination events by generating double-stranded DNA breaks in the nuclear genome facilitated by nucleases engineered to be transported during *Agrobacterium* infection, as was reported in yeast and plants, therefore making it efficient enough to be useful for biotechnology (Rolloos et al., 2015; Tzfira et al., 2003). More recently, genome editing by CRISPR has been successfully carried out in *P. tricornutum* and *C. reinhardtii*, though the latter required the use of ribonucleoproteins rather than expression of the Cas9 protein due to its toxicity (Nymark et al., 2016; Shin et al., 2016).

Despite there being only a few microalgae with established genetic toolkits, over 40 species have reportedly been transformed successfully (Gangl et al., 2015). Those in the *Chlorella* genus include *C. ellipsoidea*, *C. kessleri*, *C. sorokiniana*, *C. saccharophila*, *C. vulgaris*, *C. zofingiensis* and other *Chlorella* sp. (as reviewed in Liu and Chen, 2016; Yang et al., 2016). These were nuclear transformations using either particle bombardment, electroporation, PEG, or *Agrobacterium.* In many cases,

transformants showed only transient expression of the transgene or had problems with stability of expression (Liu and Chen, 2016; Yang et al., 2016). A large body of work has also been carried out as part of a consortium called the National Alliance for Advanced Biofuels and Bioproducts (NAABB) funded by the US Department of Energy: in addition to the sequencing of *Chlorella sp.* genomes as mentioned in **Table 1-3**, they describe the creation of a genetic toolkit and successful transformation of *C. sorokiniana* and *Auxenochlorella protothecoides* via glass-beads, though the specific details are currently unpublished and they note low transformation efficiencies of 2%-3.5% (NAABB, 2014; Unkefer et al., 2017).

Electroporation has had success in many microalgae organisms including those in the *Chlorella* genus (Liu and Chen, 2016; Yang et al., 2016). Electroporation is the use of an electric pulse to induce DNA uptake by cells. The precise mechanism is not fully elucidated but generally it is believed that the electric pulses disrupt the cellular membrane to create micropores which then allow sections of the DNA to enter the cell, whereby it is integrated in the host cell by native cellular repair mechanisms. This results in random integration into the host genome. Electroporation is another nuclear transformation technique though it has been used for chloroplast transformation in one diatom species (Reviewed in (Bashir et al., 2016)). Electroporation is a common method for use in *Chlorella* sp. having been the reported method for work on *C. ellipsoidea*, *C. saccharophila*, *C. vulgaris* and *C. zofingiensis* (Liu and Chen, 2016; Yang et al., 2016).

Additionally, *C. sorokiniana* UTEX 1230 has been transformed via *Agrobacterium* (Hiegle, 2014) (Barbi, Hiegle & Purton, unpublished data). Owing to literature reports of higher efficiencies for e.g. *Chlamydomonas sp.* than some other methods (Pratheesh et al., 2014), *Agrobacterium*-mediated transformation is the main method of choice for this project.

4.1.3 Overview of *Agrobacterium* transformation method

Agrobacterium-mediated transformation is a key method in the production of transgenic plants and was first reported for the microalga *C. reinhardtii* in 2004 (Kumar et al., 2004). *Agrobacterium tumefaciens* is a soil dwelling plant pathogen

which causes 'crown gall' disease: oncogenic tumour (gall)-like growths seen typically at the crown (above ground) -to-root junction at the base of tree-trunks and other woody plants (Kado, 2014). This pathogenicity depends on the presence of a Ti (tumour-inducing) plasmid in the bacterium, which when induced by host signals has the ability to mediate the transfer of part of its own DNA (T [transferred]-DNA) containing oncogenes and opine (nutritious amino acid derivatives) metabolism genes into the genome of the host in order to manipulate the phenotype and confer on the bacteria a competitive advantage (**Figure 4-2**).

The crux of using *Agrobacterium* as a universal transformation vector for DNA transfer to plants or other organisms lies in the manipulation of the T-DNA on the Ti-plasmid. In wild-type *Agrobacterium*, the production, transfer and integration of the oncogene-encoding T-DNA from the bacterial Ti-plasmid to the host is mediated by the host-signal induced expression products of a set of virulence (*vir*) genes which also lie in the Ti-plasmid. The T-DNA region is well defined by two inclusive direct repeat sequences of 25-28 bp termed the left and right borders, and which are required for T-DNA strand excision during infection (Gelvin, 2003). Any sequence within and including these borders is integrated into the host chromosome; therefore, replacing the native bacterial genes with target genes of interest means these will get transferred to the target (**Figure 4-2**).

Advantages of this method include the potential for transfer of large (>100 kb) DNA segments to a single locus (making metabolic pathway insertion feasible), low cost, generally single or low-copy insertion into the nuclear genome, and reportedly high efficiencies and stability compared to other methods (Gangl et al., 2015; Qin et al., 2012). There are also binary vectors available that are compatible with modern synthetic biology cloning strategies (Leclercq et al., 2015). A disadvantage is the impact on expression of positional effects or disruption of host genes as a result of random integration into the host genome, although this is a common issue with most nuclear transformation methods in microalgae (Liu et al., 2013).

4.1.4 A comparison of *Agrobacterium* transformation methods in microalgae and insight into important parts of the infection mechanism

The examination of the process of *Agrobacterium*-mediated transformation and examples of protocols in the literature allows informed choices to be made about parameters in the protocol (**Figure 4-2**, **Figure 4-3**). From the mechanism (reviewed in: (McCullen and Binns, 2006; Meyers et al., 2010; Pacurar et al., 2011; Tzfira and Citovsky, 2006)), it can be seen that crucial parts include appropriate induction of the virulence genes by chemical and physical signals, and a requirement for recognition of, and physical attachment to, the host cell (**Figure 4-2**). A survey of published protocols (detailed tables are attached in the appendix **Table A-3**, **Table A-4**, **Table A-5**) highlighted a wide variety of parameters which varied in optimal value between algal species used; a summary of these considerations is shown in **Figure 4-3**.

Although sporadic, there have been reports of over 25 species of microalgae transformed using *Agrobacterium*, starting with *C. reinhardtii* in 2004 (Kumar et al., 2004; Kumar and Rajam, 2005). In fact, *Agrobacterium*-mediated transformation has been reported in all species in the Introduction **Figure 1-1**. *Chlorella* sp. are among those successfully transformed and include *C. emersonii* (Chiaiese et al., 2011), *C. pyrenoidosa* (Reddy et al., 2017), *C. vulgaris* (Cha et al., 2012, 2011b; Ng et al., 2016) and other unspecified *Chlorella* sp. (Lin et al., 2013; Sanitha et al., 2014). One of the problems with these reports is the large amount of variation in parameters between protocols, which are often not justified or properly explained (parameters listed in appendix **Table A-3**,**Table A-4**). Also, the data on efficiencies is hard to compare because it is expressed in different ways, and stability of lines is rarely reported (appendix **Table A-5**). In fact, a recent report for *C. reinhardtii* (Mini et al., 2018) which took a thorough approach to characterising their transformant lines was unable to replicate the original Kumar *et al.* 2004 method upon which the majority of other reports are based. They had to add a modification to properly induce *Agrobacterium* virulence as discussed later.

As a starting point in this work, a protocol developed previously for *C. sorokiniana* is used (Hiegle, 2014). However, upon examination of published methods

and taking into consideration the mechanism of virulence in wild-type *Agrobacterium*, a list of important factors affecting efficiency was made (**Figure 4-3**) and modifications to this starting method gained from these insights were tested in order to optimise the method.

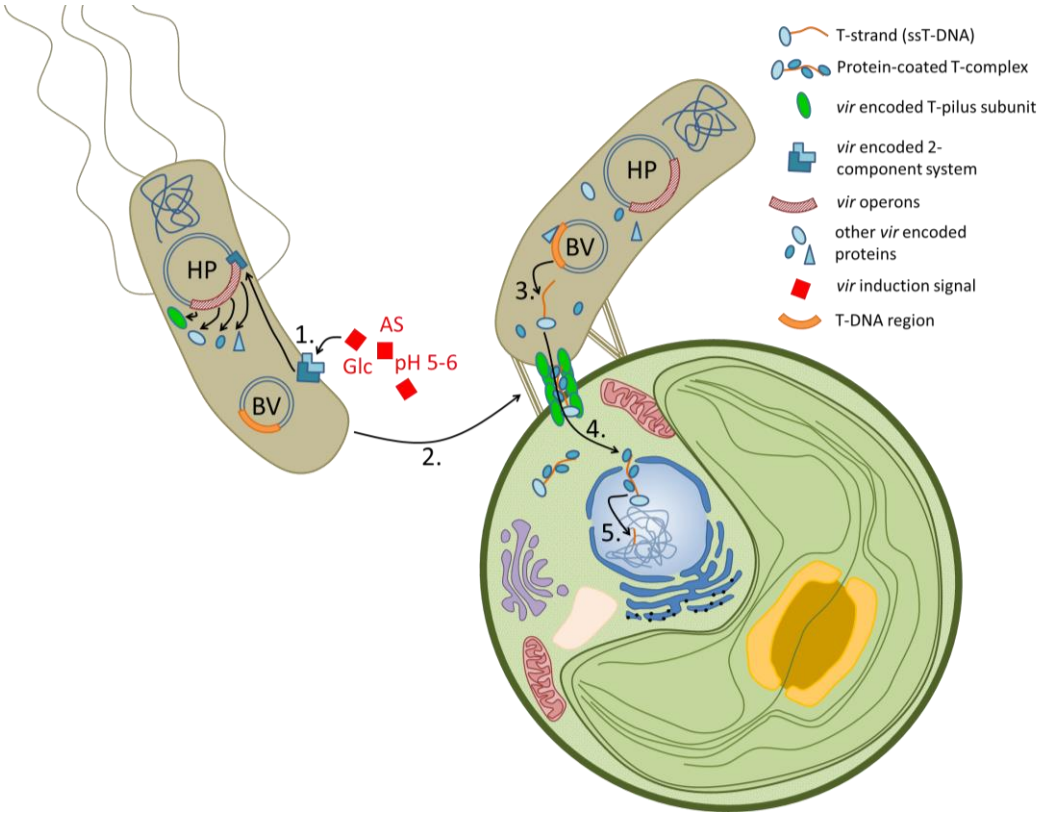


Figure 4-2 Summary of the steps involved in *Agrobacterium*-mediated transformation of microalgae. Shown is the mechanism of the transformation of a host cell nucleus using a domesticated *Agrobacterium* strain containing a disarmed helper plasmid (HP) and an appropriate binary vector (BV). **1.** Some chromosomal genes and a virulence (*vir*) gene encoded 2-component system (L-shapes) detect signals such as low pH (pH 5-6), presence of sugars (e.g. Glc, glucose) and phenolics (e.g. AS, acetosyringone). The system propagates the signal to induce various other *vir* genes/operons on the HP, which produce proteins essential for infection (see symbol legend). **2.** *Agrobacterium* attaches irreversibly to the host cell wall, causing host cell aggregation. At the cell pole, a Type-IV Secretory System T-Pilus is formed, facilitated by more *vir*-encoded products. **3.** Meanwhile, other *vir* proteins facilitate the creation of the T-strand, a ssDNA copy of the T-DNA region on the BV (highlighted segment) as defined by the left border (LB) and right border (RB) DNA sequences. **4.** *Vir* proteins guide the T-strand through the T-pilus where it becomes associated with other *vir* proteins to form a protein-coated T-complex. **5.** The T-complex moves to the host nucleus via the interaction of host and *vir*-encoded factors, where it then randomly integrates into the host genome (unless engineered otherwise).

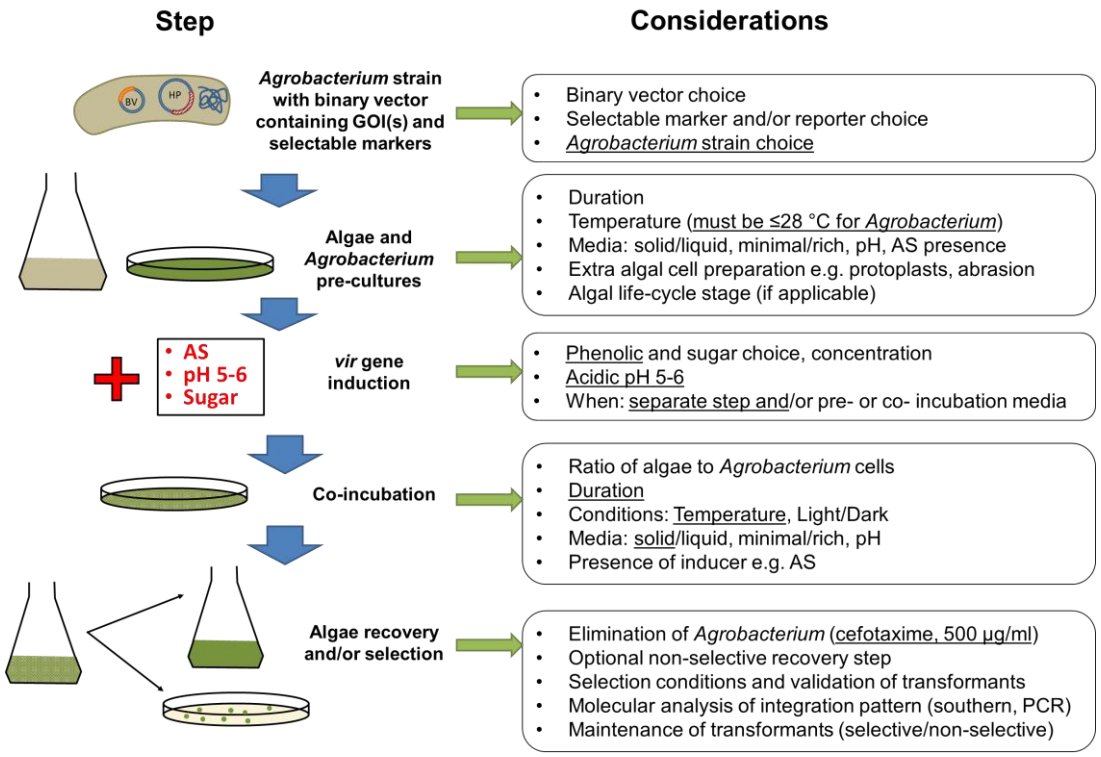


Figure 4-3 Summary of steps and points for consideration for development of *Agrobacterium*-mediated transformation of microalgae. Critical parameters are underlined. AS: acetosyringone.

4.1.5 Aims and objectives of chapter

In this chapter the aim is to find and test a suitable method for transforming the nuclear genome of *C. sorokiniana* UTEX 1230. This fits into the overall aims of the thesis because transgene expression is required to complete the omega-3 LC-PUFA pathway for proof-of-concept production of these high value oils.

This chapter describes firstly preliminary trials of electroporation and *Agrobacterium*-mediated transformation which highlighted the issue of selection problems and false positives. Subsequently, in order to improve selection efficacy, assays of wild-type antibiotic sensitivity were carried out. The chapter finishes with a focus on the preferred method of *Agrobacterium*-mediated transformation and tests modifications made to the original protocol based on the factors discussed above.

4.2 Investigations into nuclear transformation of *C. sorokiniana* by electroporation

4.2.1 Experimental parameters and conditions tested

In order to assess the transformation of *C. sorokiniana* by electroporation, the test plasmid RbleC23 (**Figure 4-1**) was used (Hiegle, 2014). The key parameters in electroporation are the electric field intensity, pulse duration, and number of pulses (Coll, 2006). Also, the electroporation medium can have an effect, especially the presence of sugars, such as sucrose, to maintain osmolarity (Coll, 2006). In this experiment the electroporation buffer used was TAP + 40 mM sucrose which is commonly used in the literature (Coll, 2006; Yamano et al., 2013) and in a commercial kit (GeneArt Chlamydomonas Engineering Kit, Invitrogen). The electroporation settings (see section 2.7.2) were partly based on literature values (reviewed in (Coll, 2006)) and those recommended for fungal electroporation instruction manual for the BioRad GenePulser XCell machine.

The linearization state of the DNA and cell concentration also vary between protocols and were the chosen parameters to test in this experiment. For DNA linearization state this was tested as circular plasmid or a single cut to linearize the whole plasmid in the backbone. In order to test cell concentration (and hence the DNA:cell ratio) the amount of DNA in the cuvette was fixed at 2 µg, but cell concentration varied from 10^6 to 10^{10} cells/ml. (**Table 4-1**).

Cell concentration (cells/ml)	**No. cells in 300 µl cuvette**
10^6	3×10^5
10^8	3×10^7
10^{10}	3×10^9

Table 4-1 Cell concentrations for testing electroporation of *C. sorokiniana*.

After electroporation and recovery, cells were plated on selective media (G418 100 µg/ml) apart from the control reactions containing no DNA where the recovered culture was split in half for plating additionally in non-selective conditions

in order to assess cell viability. This means for the control cultures there is half as many cells on each plate which needs to be considered when doing comparisons with the other conditions.

4.2.2 High level of background hinders selection of putative transformants

For the highest cell concentration (1×10^{10} cells/ml), no colonies were observed for any DNA types or on the positive control (no antibiotics) suggesting cells (**Table 4-2**) suggesting cells at this concentration are unable to survive the shock of electroporation. An observation during the electroporation process was that the cuvettes with the highest cell concentration were more likely to spark or 'arc'. In the positive control (no DNA, no antibiotics), 1400 CFU were formed when 1×10^8 cells/ml was used, and 17 CFU were formed when 1×10^6 cells/ml cuvette were used. This is a two orders of magnitude reduction in CFU from a two orders of magnitude reduction in cuvette cell density which shows that at these concentrations cell viability is similar, in contrast to the highest cell concentration where viability is zero (**Table 4-2**).

On the selective plates, there were fewer colonies than on the non-selective positive control plates, with the exception of the linear DNA condition at 10^8 cells/ml. However, the presence of large numbers of colonies on the negative controls (no DNA added to electroporation cuvette) shows the high level of background resistance to the antibiotic at the concentration of 100 µg/ml (**Table 4-2**). The only electroporation conditions which produced more colonies than the negative control is linear DNA at 10^8 cells/ml with ~2000 colonies compared to ~800 colonies on the equivalent 'no DNA' plate. The increased survival of cells from this electroporation condition suggest some of these colonies may be transformants, but due to the level of high background it was unfeasible to screen for positive transformants so no colonies from this experiment were selected for further analysis. One interesting observation is that on the selective plates, more colonies from the negative 'no-DNA' condition appear to bleach than for the DNA-containing condition. This could be indicative of varying levels of background resistance and suggests that the antibiotic takes time to kill the cells because it is an inhibitor of protein synthesis.

Selection conditions	DNA type	Cell conc. (cells/ml)	No. Colonies day 8	No. Colonies day 23
TAP agar with G418 at 100 µg/ml	Linear	1×10^{10}	0	0
		1×10^{10}	0	0
		1×10^{8}	~2000. Varying sizes, mostly small, pale.	>1200. Very varying sizes, some are much larger. About half are bleached white
		1×10^{6}	1, very small	2, very large, dark green
	Circular	1×10^{10}	0	0
		1×10^{10}	0	0
		1×10^{8}	~400. Mostly small, pale green	~500. Very varying sizes. About half are bleached white (dead).
		1×10^{6}	0	0
	No DNA	1×10^{10}	0	0
		1×10^{8}	~800, most very small, pale green	~900, very varying sizes. The majority are bleached white
		1×10^{6}	2, very small	2, very large, dark green
		No cells	0	0
TAP agar with no antibiotics	No DNA	1×10^{10}	0	0
		1×10^{8}	~1400, dark green, varying sizes	~1400. Very clear, all similar small sizes, dark green.
		1×10^{6}	17, dark green, large	17, dark green, large
		No cells	0	0

Table 4-2 Number of colonies on selective and non-selective plates from electroporation of *C. sorokiniana* in different conditions. DNA digestion and concentration of cells in the electroporation cuvette were varied. Colonies were counted using OpenCFU software (Geissmann, 2013).

4.2.3 Discussion

One advantage of electroporation is that it may be quicker than the *Agrobacterium*-mediated method (which takes over 2 weeks to complete the procedure). Stable nuclear transformation within the *Chlorella* genus has been reported previously (Bai et al., 2013; Chen et al., 2001; Chow and Tung, 1999; Gomma et al., 2015; Hsieh et al., 2012; Huang et al., 2006; Koo et al., 2013; Liu et al., 2014b; Zhang et al., 2014) though some studies report only transient expression (Maruyama et al., 1994; Niu et al., 2011; Wang et al., 2007, 2014) including in some *Chlorella sorokiniana* strains (Hawkins and Nakamura, 1999; Miller et al., 1994). However, the

protocols vary greatly in aspects such as cell growth phase, cell condition (protoplast/whole cell), cell to DNA concentration, DNA type (linear/circular), recovery steps, buffers, washes, voltage and resistance and the reasoning or full details of these parameters is not always clearly explained. It is also clear that parameters need to be optimised for each species.

A set of conditions based on the literature and other advice was used as a starting point for electroporation in *Chlorella sorokiniana* UTEX 1230 with the intention of narrowing down on optimal conditions. In the experiment the type of DNA (circular or linear) and the concentration of cells in the cuvette was tested. Although colonies were obtained on some antibiotic selection plates, this was not consistent and often equivalent numbers were also found on the negative control plates. There were also problems with cuvette arcing. Although to draw any firm conclusions this experiment would need to be repeated with optimised selection, it did potentially provide a useful starting point for future tests in that a cell concentration of 10^8 cells/ml and linear DNA were able to produce more colonies than the negative control.

In addition to the problems with selection in this experiment, another issue could have been that the cells used for electroporation had an intact cell wall. For many algal species, cells are treated with enzymes to degrade the cell wall, or cell-wall-less mutants are used. There are limited reports on those with an intact cell wall, and they require a more complicated pulse protocol (Gangl et al., 2015). If this experiment were to be repeated, the enzymatic treatment would be recommended. Additionally, the specific parameters such as pulse length, type and resistance should be assayed. Despite this method not being pursued further in this work, it is certainly much simpler and quicker than *Agrobacterium*-mediated transformation (though it requires more DNA and specialist equipment) and is worthwhile pursuing for this reason.

Another issue it that the large size of the RbleC23 plasmid used may make the procedure more inefficient (Stevens and Purton, 1997). In future work it could be beneficial to either use a smaller plasmid or to excise the specific cassette region and

use only this in the DNA mixture. This approach could also avoid unwanted integration of parts of the plasmid backbone which although may have no phenotypic effect are undesirable as they are an uncharacterised impact which could undermine the precision of genetic engineering and raise problems in industry. Carrier DNA is also sometimes used in some protocols but is not essential so it was not included here, as it may also result in some unknown integration of foreign material into the host genome.

4.3 Initial test of nuclear transformation of *Chlorella sorokiniana* using *Agrobacterium*-mediated transformation

Although electroporation may have promise as a technique, the majority of work in this chapter focuses on *Agrobacterium*-mediated transformation due its previous success in many algal species including *C. sorokiniana* (Hiegle, 2014).

4.3.1 Test construct and initial method

The test construct used in these initial *Agrobacterium*-mediated transformations was RbleC23 (**Figure 4-1**). The initial protocol developed previously is summarised in **Figure 4-12** (Hiegle, 2014). The protocol consists of co-incubating of binary-vector containing *Agrobacterium* (strain EHA 105) with algal cells on solid media for 2 days in the dark at 25 °C (**Figure 4-12**). Twenty million log phase algal cells are pre-incubated in the light at 25 °C for 1 day on solid TAP agar supplemented with acetosyringone to 100 µM. *Agrobacterium* is cultivated overnight in rich YENB medium then supplemented with 150 µM acetosyringone for 2 hours before adding 200 µl of this culture to the pre-incubated algal cells for co-incubation. After recovery of algal cells overnight in TP with 250 µg/ml carbenicillin, selection is carried out in two parallel streams: liquid selection followed by solid selection (LSS), or just direct solid selection (dSS). For LSS, 130 million recovered algal cells were incubated in a 100 ml shake flask containing 50 ml TAP supplemented with 50 µg/ml G418 and 250 µg/ml carbenicillin until the cultures start to turn obviously green in colour. At this point 1.5 ml of the selective culture is plated out using soft agar overlay onto three TAP agar plates supplemented with G418 (350-450 µg/ml), carbenicillin (500 µg/ml)

and cefotaxime (500 µg/ml). The latter two antibiotics are to eliminate residual *Agrobacterium*. For dSS, ~90 million cells from the recovery culture were plated per plate using soft agar overlay onto solid TAP containing the same antibiotics as above (**Figure 4-12**).

4.3.2 Selection conditions result in high background and a large proportion of putative transformants are lost upon subculture

A trial of the transformation was performed as non-consecutive duplicates. Selection plates were G418 concentrations of 350, 400 and 450 µg/ml (G350, G400 and G450). For the first transformation, no colonies were observed on dSS plates even after 32 days incubation. For the second transformation, dSS was not carried out. For both transformations, a LSS positive control plate containing no antibiotic was also included for both transformations and lawns were visible within 2 days.

In the first transformation, the liquid selection step of LSS greened after 10 days, but so did the negative control (negative control is co-incubation performed with wild-type *Agrobacterium*). Despite this, the cultures were still plated out and were checked at 7 and 22 days. There was a lawn after 7 days on all RbleC23 plates though on the G450 plate this was restricted to one edge of the plate. At 22 days large, dark green colonies were visible on RbleC23 G450 and the negative control (only plated at G350), though they were less numerous compared to RbleC23. No large colonies were visible on the G350 or G400 RbleC23 plates, but the lawns were still visible though they appeared slightly faded in colour indicating some bleaching and cell death.

For the second transformation, half the number of cells were added to the liquid selection step of LSS (65 million cells per flask instead of 130 million cells). The liquid cultures greened within 5 days and the subsequent LSS plates were checked after 20 days. A partial lawn (not covering the whole plate) was visible on G350 concentration. On G400 and G450 hundreds of dark green single colonies of varying sizes were visible but with fewer on the G450 plate.

Colonies from LSS plates of both transformations were picked for further propagation on G450 plates with 500 µg/ml carbenicillin to continue to ensure *Agrobacterium* had been eliminated. 75-78 % of picked colonies died upon the first propagation round but after that they all survived (**Table 4-3**). Of those that survived, all were from the G450 plate except for one from the G350 plate and one from the negative control plate of the first transformation. This suggests that at the lower concentrations of the antibiotic there could have been spontaneous resistance to that concentration only, or if they were real transformants then expression was very low.

After three rounds of selective propagation, putative transformants were maintained on plain TAP with no antibiotics. Phenotypic tests on both expression cassettes were performed after the third propagation had grown and are described in the next section.

Transfor-mation	**Number of colonies picked and from which plates**				**Survival propagation 1**				**Survival propagation 2 onwards**
	G350	**G400**	**G450**	**-ve**	**G350**	**G400**	**G450**	**-ve**	**All**
First	3	3	28	2	1	0	7	1	Unchanged
Second	3	3	32	0	0	0	8	0	Unchanged

Table 4-3 Survival of colonies picked from selective plates of a test of *Agrobacterium*-mediated transformation of *Chlorella sorokiniana* UTEX1230. Selection was on G418 at 350, 400, and 450 µg/ml and propagation was at 450 µg/ml.

4.3.3 Spot tests show that putative transformants have increased resistance to the reporter antibiotic, though an escape mutant also thrives

Spot tests were performed using cells grown from the third propagation plate after 14 days incubation. Spot tests were performed both on the selectable marker G418, and the reporter marker zeocin. Cells were not exposed to the reporter marker zeocin before this, so had faced no previous pressure to form resistance. Using a pipette tip, a small amount of cells were inoculated into 1 ml non-selective TAP medium in a 2 ml tube and incubated at 24 hours in the light at 25 °C, 120 rpm. Three five-fold serial dilutions were conducted and 10 µl of each dilution plus 10 µl

undiluted culture was spotted onto plates containing either no antibiotics, G418 at 450 µg/ml, Zeocin™ at 250 µg/ml or Zeocin™ at 350 µg/ml. Plates were incubated in the light at 25 °C for 13 days before being photographed.

All cells grew on the no antibiotic control but the cell concentrations are noticeably smaller for some colonies, highlighting one disadvantage of this method and the importance of comparing the results to the positive control. The wild-type also grew to higher cell density than any of the putative transformants, a lawn forming even at the most dilute spot compared to individual colonies for the others (**Figure 4-4**, first column "TAP").

All putative transformants had more G418 resistance than the wild-type, which did not grow on any plate (**Figure 4-4**, G450). However, the colony isolated from the negative control transformation plate also grew at this G418 concentration suggesting spontaneous resistance is possible. Though as seen in the fourth dilution, the growth is attenuated compared to the positive control on TAP without antibiotics. Putative transformants which appear not to show attenuated growth on G450 include 1a, 2c, 1b, 1c, 1e, 1f and 1g (six from the first transformation including one from the G350 plate, and one from the second transformation).

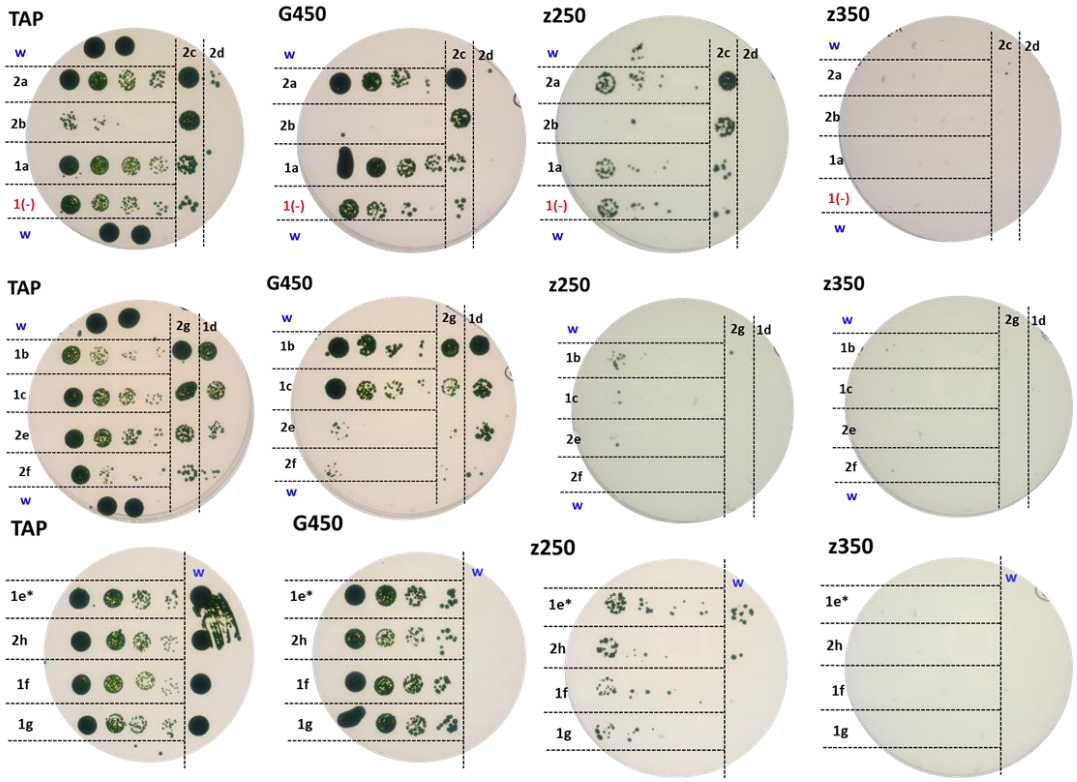


Figure 4-4 Spot tests of putative transformants of *Chlorella sorokiniana* UTEX 1230 using G418 and zeocin. The spots were conducted as four 5× serial dilutions in left to right or top to bottom direction on each plate. The wild-type was spotted on each plate ("w", in blue). A single colony isolated from a negative control plate of one transformation was included ("1(-)", in red). The other spots are labelled according which transformation they were from (1 or 2) followed by an ID letter (a-h). The first plate (TAP) is the positive control containing no antibiotics. The other plates are G418 at 450 µg/ml (G450), and zeocin at 250 µg/ml or 350 µg/ml (z250 and z350 respectively). All putative transformants were isolated by selection of colonies from G418 and 450 µg/ml except * which was isolated from a plate at G418 at 350 µg/ml. All putative transformants and the negative isolate have survived propagation for 3 rounds on G418 at 450 µg/ml.

None of the cells had been exposed to zeocin prior to these spot tests so resistance could not have accumulated over multiple rounds of streaking. At the highest two serial dilutions for the wild type, and the highest three dilutions for the negative colony, growth was seen on zeocin at 250 µg/ml suggesting that this concentration of zeocin is not an adequate concentration for testing the expression of the *ble* cassette as it cannot kill the wild-type, or there were too many cells per

plate (see section 4.4). However, as seen in the dilutions, the wild-type was of much higher cell concentration as the spot is a lawn on the no antibiotic plate compared to single colonies with the other transformants, and in this particular spot test cell number was not strictly controlled. It could be that the appearance of colonies on the zeocin in wild-type is due to the higher cell concentration. Though this would not explain growth of the negative isolate colony though since cell concentrations were comparable to the others. Of the putative transformants listed above whose growth was not attenuated on G418, only 2c and 1b show growth on zeocin at 350 µg/ml, and this is only small colonies from the least dilute spot. Transformant 2c in particular also performs well at the z250 concentration with almost no attenuation of growth compared to the other putative transformants.

4.4 Antibiotic sensitivity tests for optimisation of selection conditions

In order to reduce background spontaneous resistance colonies or physical survivors (where the top cells on a colony are exposed to less antibiotic than those on the bottom) which were prevalent in the test transformations, tests were done to assess the adequate concentration of G418 to reliably kill the wild type. Factors which may affect this were tested including seed growth phase, soft agar use, cell number and the use of cells pre-incubated on solid agar rather than liquid culture. For zeocin, optimisation of the condition sufficient to kill the wild-type in phenotypic spot tests was assessed.

4.4.1 G418

4.4.1.1 Effect of different seed quality

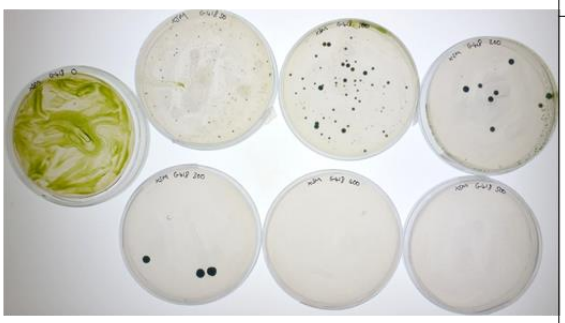
In the following experiment cells from seed cultures in two different growth phases, late log and late stationary phase, were used. Colonies appeared in both seeds up to G300 after >30 days incubation (**Figure 4-5**A). Although the late stationary phase had more colonies on the G200 plate, it had none on G300 like the late log plates, and there are no obvious differences between these two conditions. There was a visible bleaching of the lawns for all antibiotic concentrations. Colonies which do appear tend to appear in areas where there is increased cell number due to

uneven plating. In the second test a late log phase culture was plated between 300-400 µg/ml G418 and no growth was observed on any plates (**Figure 4-5**B). It was decided to use 350 µg/ml in further tests because at G300 some colonies appeared but at G400 some low expressing putative transformants may not be detected.

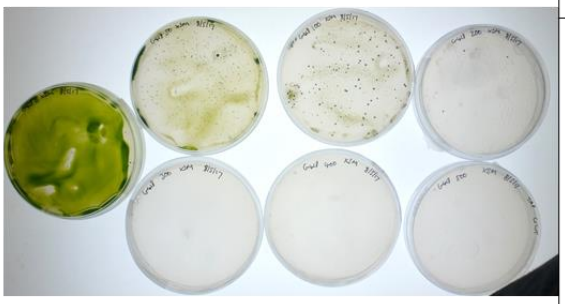
Figure 4-5 Growth of wild-type *Chlorella sorokiniana* UTEX1230 from different growth phases on increasing concentrations of G418 antibiotic. (See figure opposite/ahead). A: Concentrations top row left to right are 0, 50, 100, 200, bottom row left to right are 300, 400 and 500 µg/ml. B: Concentrations top row left to right are 0, 300, 325, bottom row left to right are 350, 375 and 400 µg/ml. 100 million cells were plated from either a late log phase of late stationary phase liquid culture.

A

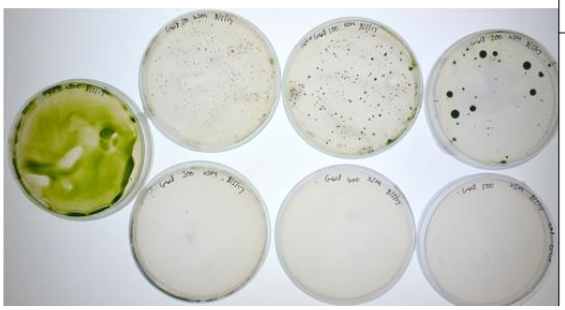
Incubation (days)	Seed type	Growth
10	Late log	Colonies visible up to G300, smaller at higher concentrations



Incubation (days)	Seed type	Growth
34	Late log	Colonies visible up to G300, large. Small and large at G100.

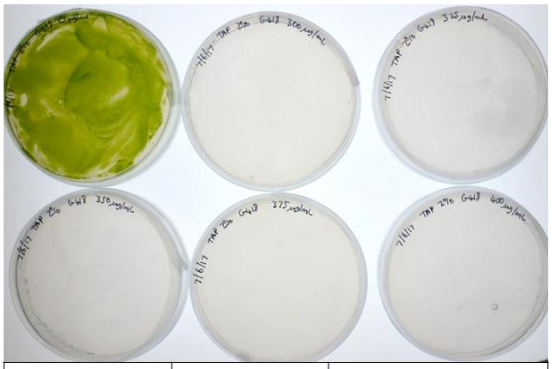


Incubation (days)	Seed type	Growth
8	Late stationary	Colonies visible up to G200, small



Incubation (days)	Seed type	Growth
31	Late stationary	Colonies visible up to G300, small. Small and large at G200.

B



Incubation (days)	Seed type	Growth
19	Late log	No colonies visible on any concentration

4.4.1.2 The efficacy of the antibiotic is affected by the number of cells plated.

In order to efficiently test a range of cell number to antibiotic ratios, serially diluted spot tests on G350 were also carried out. The seed culture was a stationary phase liquid culture and was concentrated so that an undiluted 10 µl spot contained 7 million cells. It appears that across all cell concentrations no growth is seen after 5 days (**Figure 4-6**) and that G418 at 350 µg/ml is a suitable concentration to use.

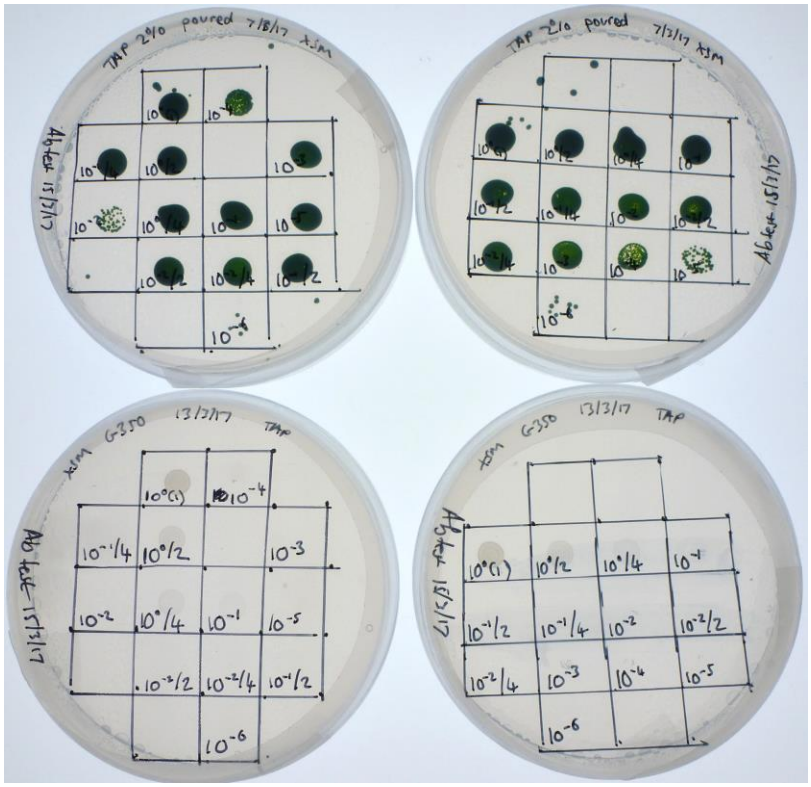


Figure 4-6 Spot tests of wild-type *Chlorella sorokiniana* UTEX1230 on G418 at 350 µg/ml after 5 days growth. The spot order on each plate was randomised for the two replicates (left and right). The top plate in each replicate is TAP agar with no antibiotics as a positive control. The bottom plate contains G418 at 350 µg/ml. The wild-type culture was serially diluted by a factor of ten, and some concentrations were also diluted by two and four as specified.

Different cell numbers were also tested on a whole plate. As discussed later, plating higher cell numbers could avoid having to discard parts of the co-incubated transformation culture, of which there are many cells. However, there is a balance between plating more cells and the level of antibiotic needed becoming unrealistic. The concentration used here was 350 μg/ml as it was determined to be adequate over a large cell concentration in the spot tests above.

For the whole plate tests, 50 million and 250 million cells were plated over different G418 concentrations. The seed cells were incubated on solid media to ensure the seed train used was the same as that used in transformation experiments rather than using liquid seed cultures since during *Agrobacterium*-mediated transformation, the *C. sorokiniana* cells are grown on solid media during co-incubation. After three days growth, no single colonies are visible for either concentration, instead there is a lawn of cells which has visibly bleached (an indicator of death) at both cell concentrations but more pronounced in the 50 million cells plate. Additionally, the bleaching is more pronounced at higher concentrations of the antibiotic (**Figure 4-7**). By the sixth day of incubation, the lawns have died back enough to allow single colonies to be visible. For the 200 million cells plate, large numbers of colonies are visible up to G300, and there is a single colony in a dense spot on the G350 plate. For the 50 million cells plated there are numerous colonies on the G200 plate, and some barely visible colonies on the G250 plate. There are no cells at the higher concentrations. After 13 days incubation the colonies are much more pronounced at both cell concentrations. For the 200 million cells per plate there are colonies up to G350 concentration suggesting that this is too many cells to plate. For the 50 million cells plate, there are only colonies up to G250 concentration, with none on G300 or G350 **Figure 4-7**. Using OpenCFU colony counting software, there were ~600 colonies at G250 for 250 million cells but only 30 for 50 million cells. For only five times more cells per plate there is 20 times more colonies at the same antibiotic concentration, showing that the relationship between cell number and antibiotic concentration is not linear in this case.

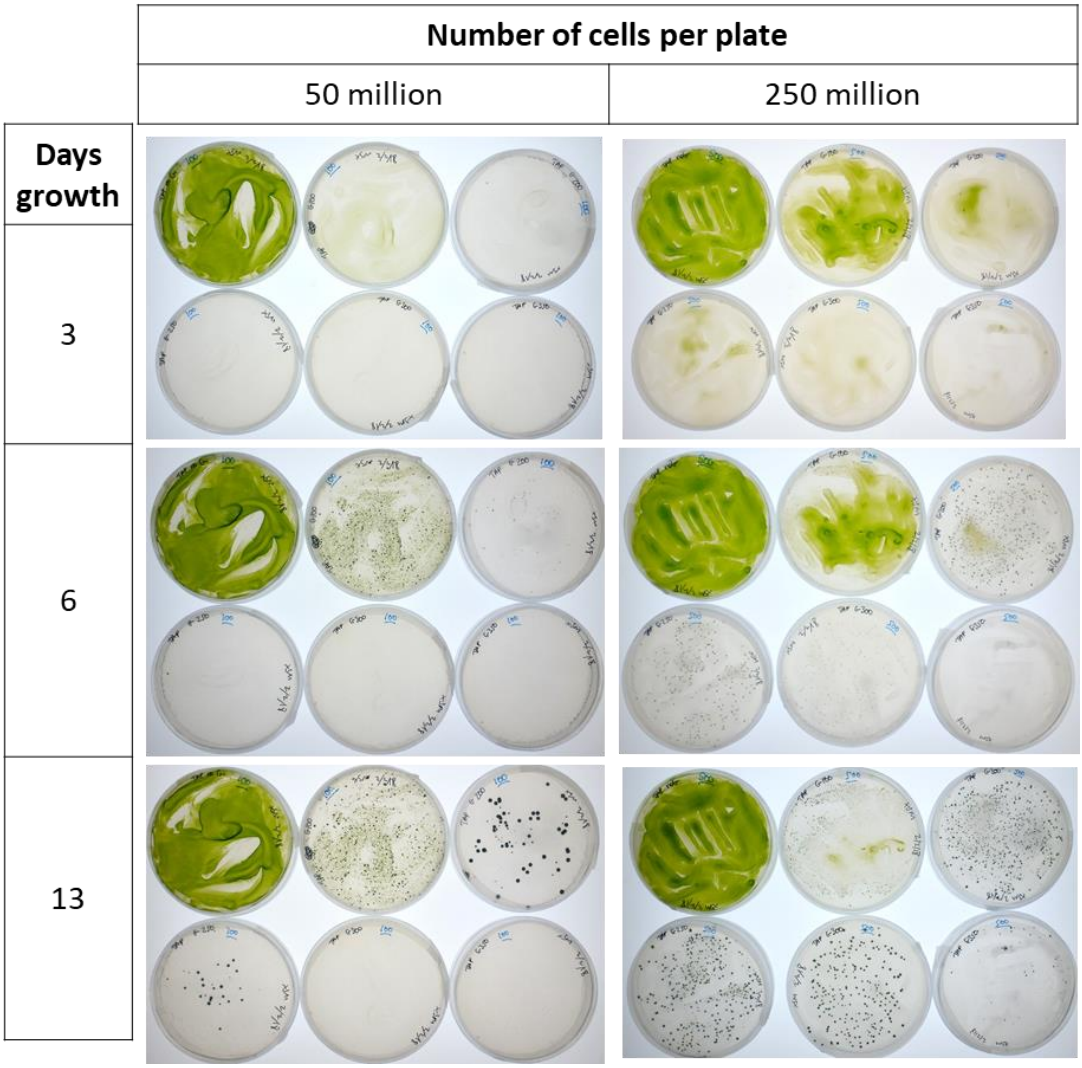


Figure 4-7 Growth of wild-type *Chlorella sorokiniana* UTEX 1230 over time plated at two different concentrations on G418. The G418 concentrations top row from left to right are 0, 100, 200, and bottom row 250, 300 and 350 µg/ml respectively.

4.4.1.3 Soft agar plating reduces spontaneous colonies

As a result of colonies occurring in areas of high cell concentration when plated using a spreader (for example **Figure 4-7**), the use of soft agar was tested to allow for more even distribution of cells and to avoid areas on the plate of particularly high cell density. The effect of soft agar was tested using 100 million cells per plate since 50 million cells in the above test was successfully killed and it is desirable to select the most cells at once from the transformation method as discussed later.

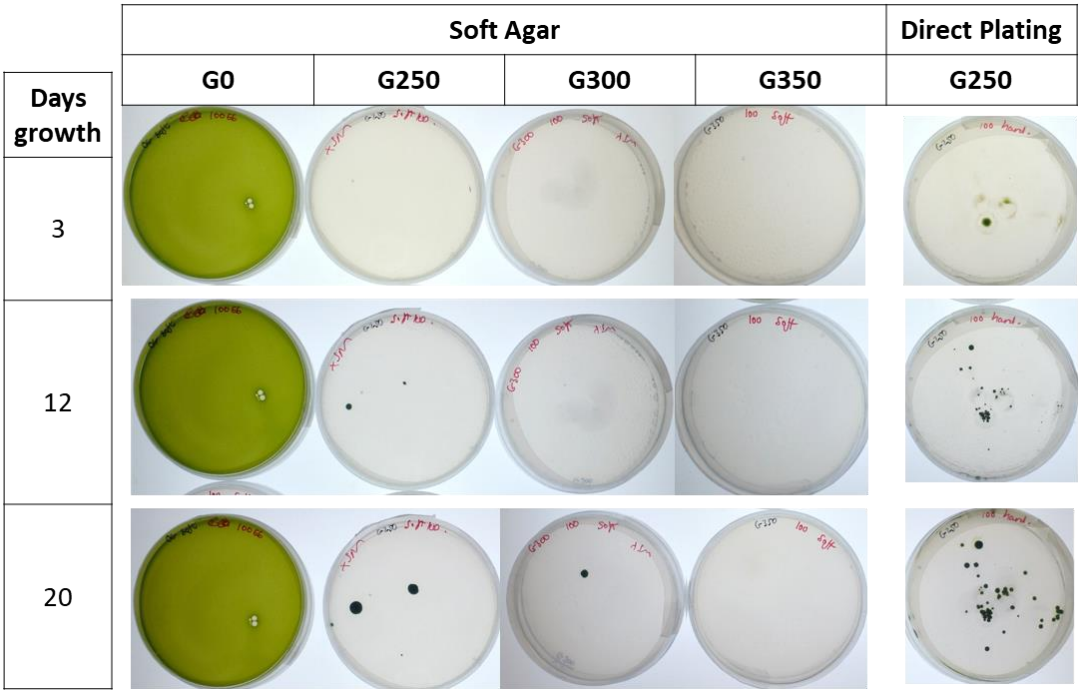


Figure 4-8 *C. sorokiniana* UTEX 1230 wild-type growth over time on G418 antibiotic plates on soft and hard agar. There were 100 million cells plated per plate.

The only antibiotic concentration which did not produce spontaneous mutants was G350 (**Figure 4-8**). For the other conditions, these mutant colonies appeared only after 12 days, suggesting that the antibiotic may be becoming less effective over time or that it is just retarding growth. Therefore, it is better to ensure colonies are picked before this time in transformations. When comparing direct plating vs soft agar at G250, it is clear that direct plating results in many more spontaneous colonies from a much earlier time, and it can be seen that these occur mostly in areas of the plate where cell concentration was highest due to uneven spreading with a plate spreader which encourages physical survivors. Therefore, it is best to use soft agar because it enables cells to be distributed evenly and encapsulates the whole cell.

4.4.2 Determining an appropriate Zeocin concentration to eliminate wild-type *C. sorokiniana*

4.4.2.1 Spot tests

Zeocin is used as the reporter antibiotic and is not used for direct selection but knowledge of the appropriate level required to kill wild-type *C. sorokiniana* cells is still needed since this is how it would be assayed in putative transformants. In this test, the dilutions for each spot were calculated so that they corresponded to different total plating densities, starting with the most concentrated spot being plated from a cell suspension at a concentration of 2.5×10^9 cells/ml (2.5×10^7 cells per 10 µl spot) which is equivalent to 5×10^8 cells per 200 µl plating volume. The monitoring of spots on different concentrations of zeocin over time is shown in **Figure 4-9**. These tests show that the spots take several days to bleach, but any colonies are visible by day 9. Zeocin at a concentration of 300 µg/ml is adequate to kill wild-type cells, as seen by the bleaching of the spots compared to the no zeocin control plate. A single colony appearing on the third spot of the top row in z300 day 9 is due to a human error. Additionally, there is a dark ring around the highest cell density spot which may be cells surviving and shows that the lower plating densities are more suitable. All other zeocin concentrations show several spontaneous colonies occurring in the spots at all cell densities so would be unsuitable in these tests.

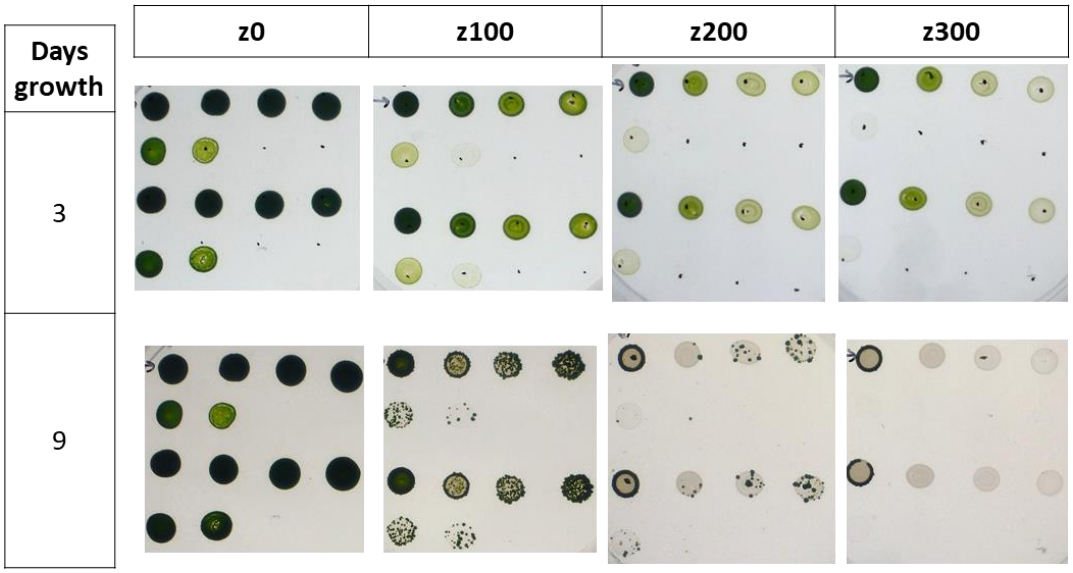


Figure 4-9 Spot tests of wild type *C. sorokiniana* UTEX1230 on different concentrations of antibiotic zeocin. The two rows of spots were duplicated for each plate to give a total of four rows. The number of cells per 10 µl spot for the first row going left to right are 25×10^6, 10×10^6, 5×10^6, and 2.5×10^6. For the second row of spots it is 10× serial dilutions starting at 2.5×10^5 cells per spot. For the image from three days, there are black dots indicating the approximate middle of the spot.

4.4.3 Testing zeocin and G418 in liquid media

Liquid selection was a part of the initial transformation protocol devised in the lab (Hiegle, 2014), and although this step was removed in the modified method, it is still useful to confirm the antibiotic concentration required to stop wild-type growth. Since some antibiotics are sensitive to light, this test was performed in both the light and the dark. To test multiple concentrations at once the test was performed in 12 well plates inoculated with equal numbers of cells from a suspension made directly from cells of an agar plate stock. In each 1 ml well there were 2.13×10^5 cells and they were incubated with shaking at 120 rpm in the light.

The plates were monitored by imaging them over time (**Figure 4-10**) and although the cultures appear paler in the dark incubation plate, the pattern of growth shows no observable differences between light and dark suggesting in this case light was not affecting the performance of the antibiotic. Zeocin prohibited growth of *C. sorokiniana* only at a concentration of 50 µg/ml until 6 days when the wells of both

light and dark began to show growth. This suggests a higher concentration is needed as cells may be able to persist at a minimal level until the level of antibiotic has degraded to a low enough concentration or is indicative that at this concentration the antibiotic only slows growth. Alternatively, some single cells may have undergone adaptation and hence have a longer lag phase since their starting concentration would be low. For G418, 30 µg/ml was adequate to prevent growth over the 11 days. At 10 µg/ml, growth was only prevented up to the tenth day, presumably because again the cells were able to persist minimally until the antibiotic had degraded enough for them to grow. This selection in 12 well plates could be a useful tool for growing multiple transformants at once and assaying the strength of their antibiotic resistance if comparisons between different promoters are to be done.

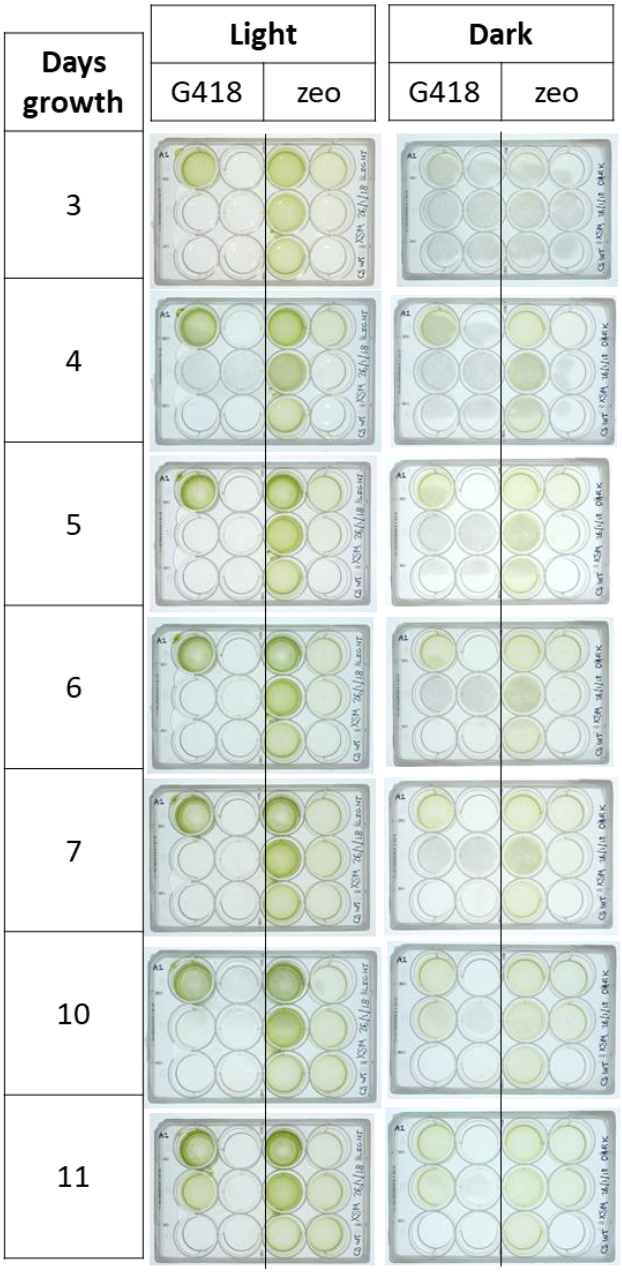


Plate Layout

G418		zeo	
None	G30	Z5	Z30
G10	G40	Z10	Z40
G60	G50	Z20	Z50

Figure 4-10 Growth of wild-type *C. sorokiniana* in small-scale liquid media containing antibiotics G418 or zeocin. Concentrations are shown. Equal numbers of a cells from a suspension made straight from an agar plate stock were added to each well.

4.5 Testing a colony PCR method for *Chlorella sorokiniana*

A common method for PCR testing of algae includes boiling a large loop of cells from a plate in Chelex® 100 ion-exchange resin (BioRad). This resin can chelate polyvalent metal ions such as magnesium, which are cofactors of DNAses, and the supernatant can then be used in a PCR reaction (see materials and methods). The

manufacturer manual for Chelex 100 notes its use for removing trace metal contaminants from biological systems whilst leaving enzymes and other proteins unaffected. Chelex 100 has been shown to perform better in both PCR reactions and storage than other lysis buffers such as TE, SDS, EDTA and Triton in multiple microalgae including *Chlorella sorokiniana* UTEX 1230 (Wan et al., 2011). However, although the Chelex-based method is relatively simple, this protocol requires multiple steps, such as centrifugation and extraction of supernatants, which are laborious and take up large amounts of time if multiple colonies need to be screened. Liu *et al*. (Liu et al., 2014a) demonstrated a protocol in several algae including *Chlorella* sp. which did not require several steps and could be performed in a single tube by boiling a colony, or cells from liquid culture, directly in the PCR tube in PCR buffer. Therefore, it was useful to try this protocol with *Chlorella sorokiniana* UTEX 1230.

Boiling time is important in colony PCR reactions, it needs to be long enough to facilitate cell lysis but over-boiling could cause the DNA to shear. The length of time needed depends on the strength of the cell wall and is variable for different *Chlorella* sp*.,* one study found that boiling was absolutely necessary for colony PCR in *C. vulgaris* but not in others (Packeiser et al., 2013). Liu *et al*. (Liu et al., 2014a) also found in their protocol that the required boiling time varied in the microalgal species tested between 5 and 20 minutes, with no PCR product at 0 min. Of the microalgal strains tested, three were of the *Chlorella* genus (none were *C. sorokiniana*) and needed different boiling times of 5 or 10 minutes. For the Chelex method, Wan *et al*. (Wan et al., 2011) found boiling at 10 min was adequate for *C. sorokiniana* UTEX 1230, which differs from the 5 minutes used in our lab protocol (see materials and methods). Due to differences between strains, and that *C. sorokiniana* was not one of those tested by Liu *et al*., (Liu et al., 2014a) it was decided to try different boiling times for *C. sorokiniana* using the single tube protocol.

Another aspect which can affect a colony PCR procedure is cell number. Colony sizes can differ greatly, as can the human error in picking colonies off plates. The authors Wan *et al*. 2011 tested algal cell numbers and found the Chelex-based method to be stable over a range of 5×10^5 to 1.5×10^7 cells being lysed in 50 µl

buffer with 1 µl of reaction product typically being used for PCR. Liu *et al*. showed that the PCR reaction worked with a minimum of 5×10^3 cells in their single tube protocol, but was best with 4×10^4 to 5×10^5.

Since cell number is important, the number of *C. sorokiniana* cells picked when using a tip was measured by touching the tip to a plate, resuspending the cells on the tip in 40 µl TAP and counting the cells. This was performed three times with different techniques; using a 10 ul tip touching a colony, using a 10 ul tip on a well-grown part of the plate, and using a 20 ul tip touching a well grown part of the plate. After resuspending in 40 ul the cell count was 26.8×10^6, 19.8×10^6 and 13.7×10^6 cells/ml respectively which gives a mean of $20.1 \times 10^6 \pm 6.56 \times 10^6$ cells/ml. This corresponds to an average of 8×10^5 on the tip (total number of cells in 40 ul). During a transformation, there will probably be fewer cells than this because the primary selection colonies are often very small.

To test the PCR reaction, the internal transcribed spacer ITS-2 of the small and large rRNA genes which is used in phylogenies and identification of microalgal strains was amplified (Hall et al., 2010; Liu et al., 2014a). The Chelex protocol described in the Materials and Methods was used as a control, and a band was obtained at ~ 750 bp (**Figure 4-11**), as expected since the product should be between 0.6-1kb (Liu et al., 2014a).

To test the single tube method, 1 or 2 ul of the resuspended culture was used as template, which should correspond to ~2.68×10^4 or 5.36×10^4 cells respectively and is within the working range shown by Liu *et al*. (Liu et al., 2014a). The PCR reaction worked best with the larger cell number, with only a faint band visible at the lower cell number after 10 minutes boiling time and no band after 20 minutes boiling time (**Figure 4-11**). Interestingly, although all studies specify that boiling is essential for *Chlorella* sp. as discussed above, it appears that no pre-boiling was necessary here, as the unboiled sample gave the strongest band (**Figure 4-11**). It could be that the 30 seconds initial denaturation at the start of the reaction was enough to lyse the cells and that at higher temperatures the DNA sheared. No PCR band is observed after boiling after 10 or 30 minutes, however a band at the same size is seen after 20

minutes boiling. Despite the unclear results with boiling time, it is clear that this single tube protocol from Liu *et al* (Liu et al., 2014a) is a suitable method for use with *C. sorokiniana* UTEX 1230 and could be a useful tool for screening high numbers of colonies. However, it should be noted that the primers used target the ITS-2, of which there are multiple copies. Therefore, more template may be needed for transformants where there is likely to be just a single copy of a transgene.

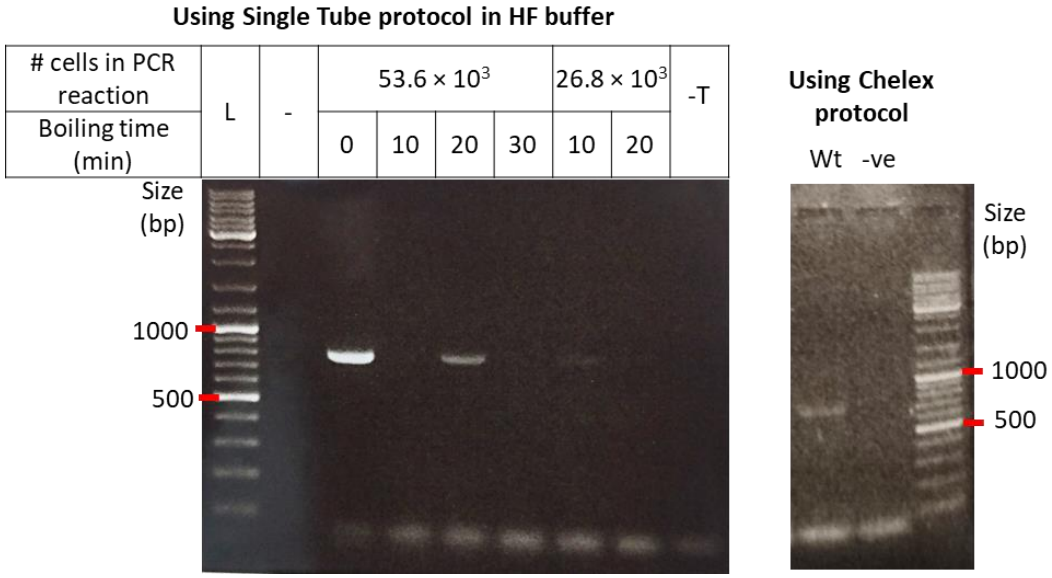


Figure 4-11 Testing different numbers of cells and boiling times on a PCR protocol for *Chlorella sorokiniana*. L: Ladder (Thermo GeneRuler DNA Ladder Mix). -T: no template.

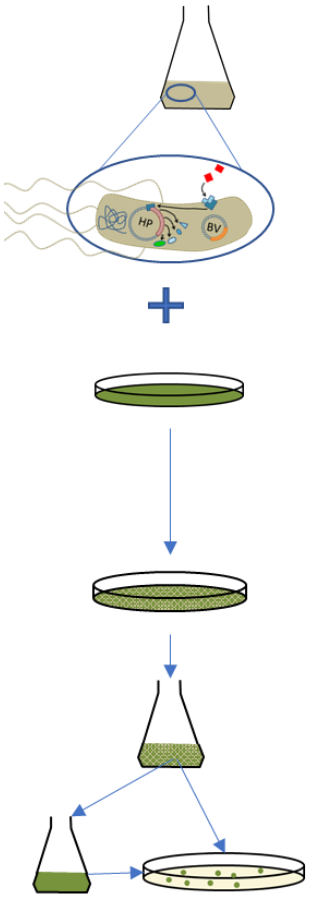
4.6 Testing modifications to the *Agrobacterium* transformation method in an attempt to improve efficiency

The large loss in putative transformants in test transformations and high background could have been due to a combination of inadequate antibiotic selection conditions and the transformation method being inefficient. In order to improve the likelihood of obtaining real transformant colonies, efforts were made to explore which parts of the method could be improved to increase the efficiency. An examination of the literature showed many aspects of the protocol which can affect the outcome (see introduction, **Figure 4-3** and appendix tables) including pre-culturing conditions, induction of virulence in *Agrobacterium*, co-culturing conditions and algal recovery. In addition, the number of algal cells after co-incubation must be

considered, as there is a limit to the cell number at the optimised antibiotic concentration as shown in section 4.4. Therefore, if there are too many algal cells at the end of the procedure then it is impractical to plate them all out and some culture is discarded which means true efficiency is not obtained.

The modifications made to the method tested are discussed in turn below and include (**Figure 4-12**):

1. A decrease in the number of algal cells plated out in pre-incubation and the removal of the recovery step before selection
2. Alternate growth conditions for *Agrobacterium* to further enhance virulence induction
3. Altering the ratio of *Agrobacterium* to algal cells during co-incubation
4. Performing direct solid selection only (removal of the liquid selection)



Step	Original method	Modifications
Agrobacterium pre-culture	Overnight 28 °C in YENB (rich media)	Unchanged
Agrobacterium virulence induction	Addition of 100 µM acetosyringone to pre-incubation culture two hours before coincubation	1 day at room temperature in AB minimal medium pH 5.5 containing 0.5% glucose and 100 µM acetosyringone
Algae pre-culture	20 million cells from log-phase liquid culture plated on solid TAP agar containing 150 µM acetosyringone 1 day, light, 25 °C	Unchanged except 5 million cells
Co-incubation	Agrobacterium OD 600 =1 Dark, 2 days, 25 °C	Unchanged but Agrobacterium OD 600 =5
Recovery	Overnight in liquid TP medium, light, 25 °C, shaking	Step not included
Selection	1. Direct solid selection 2. Liquid selection then solid selection	Direct solid selection only

Figure 4-12 Summary of initial and modified *Agrobacterium*-mediated transformation method. The initial method is from (Hiegle, 2014). Justification for the modifications is discussed in the text.

4.6.1 Reducing the number of algal cells so it is practical for selection

As tested in the antibiotic sensitivity section (4.4), there is a finite number of cells that can be plated per selective G418 plate otherwise there is increased background, and from the conditions tested this was 100 million cells on a concentration of 350 µg/ml. In the original protocol, by the end of the co-incubation and overnight recovery, there were so many algal cells that a large portion of the culture had to be discarded before selection because it is impractical to plate it all out, hence many putative transformants are potentially lost. In one run of the original protocol, after recovery there was a total of 1.8×10^9 cells. Of these cells, 130 million

were used for liquid selection in 50 ml of medium, of which 1.5 ml is plated out for colony isolation. This means only ~7 % of the recovery culture was used for liquid selection, and of this liquid culture only 3 % was plated out. For direct solid selection, ~90 million cells were used per plate which only represents 5 % of the recovery culture. Selection tests on all the recovery culture would be impractical. For example, using the above numbers as an example, it would require a total of 13 liquid selection cultures, each requiring 30 plates for colony isolation. For direct solid selection, a total of 20 plates would be needed. To scale this up, for example if testing a library of parts, then this rapidly becomes unfeasible and in reality, only one or two selective plates per transformation are used, which means only ~5 % of the original co-incubation culture is selected, and the rest is discarded.

Therefore, reducing the number of algal cells by the end of the procedure would be beneficial to reduce the number of false positives. This was considered in two ways, by reducing the initial number of algal cells at pre-incubation, and removal of the overnight recovery step as detailed below. The number of algal cells plated on the pre-incubation plate was reduced from 20 million to 5 million. 5 million cells at pre-incubation was used successfully in a published *C. vulgaris* transformation using *Agrobacterium* (Cha et al., 2012; Sharif et al., 2015). The recovery step was removed because it unnecessarily increased the cell number, despite using the autotrophic medium TP in which *C. sorokiniana* grows slowly. Also, the purpose of a recovery is to allow cells to begin to express the transformed genes and accumulate the transgenic product, but the algal cells have time over the course of the co-incubation to begin this expression. In the literature it has been found that the recovery step is not needed, and in some cases even reduced transformation efficiency (Dautor et al., 2014; Ubeda-Minguez et al., 2015).

The effect of reducing the number of algal cells at pre-incubation and removing the overnight recovery step was measured by counting cells at various stages during the protocol. From the 5 million initial cells plated, there were 46 million cells after pre-incubation, which is an increase of 9.2-fold. After co-incubation, a mean cell-count of three plates was 1024 million +/- 85 million (standard deviation) which is a further increase of ~22-fold. By plating 100 million cells on selective media,

each plate contains ~10 % of the population, double the amount from the original protocol. However, even with these modifications it still requires ten plates per construct which is still not ideal for multiple transformations. To further increase the population of cells able to be plated out, more concentrated antibiotic selection conditions would need to be tested.

4.6.2 Transformations with extra *Agrobacterium* virulence induction steps

Virulence induction is a crucial part of the *Agrobacterium* infection cycle where the cell prepares itself to undergo host attachment and build the components required for transfer and integration of the T-DNA in the host nucleus, and it is tightly regulated due to the considerable effort required by the cell (see section 4.1.4 and **Figure 4-2**). Therefore, adequate induction is important to consider in microalgal transformation protocols.

Many microalgal protocols simply use supplementation of *Agrobacterium* or algal growth medium with the phenolic compound acetosyringone, typically at 100-150 µM. However, the use of acetosyringone alone without other synergistic factors such as acidic pH of 5.5, is supposedly ineffective and a common mistake even in some plant protocols (Gelvin, 2006). Indeed, examination of the *Agrobacterium* infection mechanism (see introduction) shows that multiple induction factors are required. Also, a recent study involving *C. reinhardtii* (Mini et al., 2018) found that they could not replicate the original version of the *Agrobacterium*-mediated protocol for *C. reinhardtii* (Kumar et al., 2004), which does not contain a separate virulence induction step. They were also unable to replicate a modified version with reportedly higher efficiency which included an induction step for 4 hours (Pratheesh et al., 2014). Instead Mini *et al* 2018 had to develop a new protocol with a much longer virulence induction period based on Gelvin's report (Gelvin, 2006). Since the initial protocol only partially included virulence induction by the apparently ineffective strategy of adding acetosytringone to rich media, it was decided to introduce a modification based on both Mini *et al* 2018 and Pratheesh *et al* 2014. The difference in virulence induction is summarised in **Figure 4-12**.

The new virulence induction method involved pre-incubation of cells in rich medium followed by resuspension, and incubation for 24 hours in acidic minimal AB medium containing acetosyringone at room temperature rather than 28 °C which is supposed to also aid virulence induction (Gelvin, 2006; Wise et al., 2006). Due to very slow cell growth in the induced cultures overnight (**Figure 4-13**), the induced cultures were sub-cultured into new rich media overnight and the OD at 600 nm was monitored for an increase as an indicator of cell viability, since cells must be alive to infect the host organism. In this overnight growth experiment, the sub-cultured, induced cells grew as well as the positive control which was a single colony inoculated into the same growth medium. These cells were grown in the presence of 50 µg/ml kanamycin; resistance to which is conferred by the resistance gene on the binary vector. This confirms that this vector (which also contains the T-DNA), is still present in the cells.

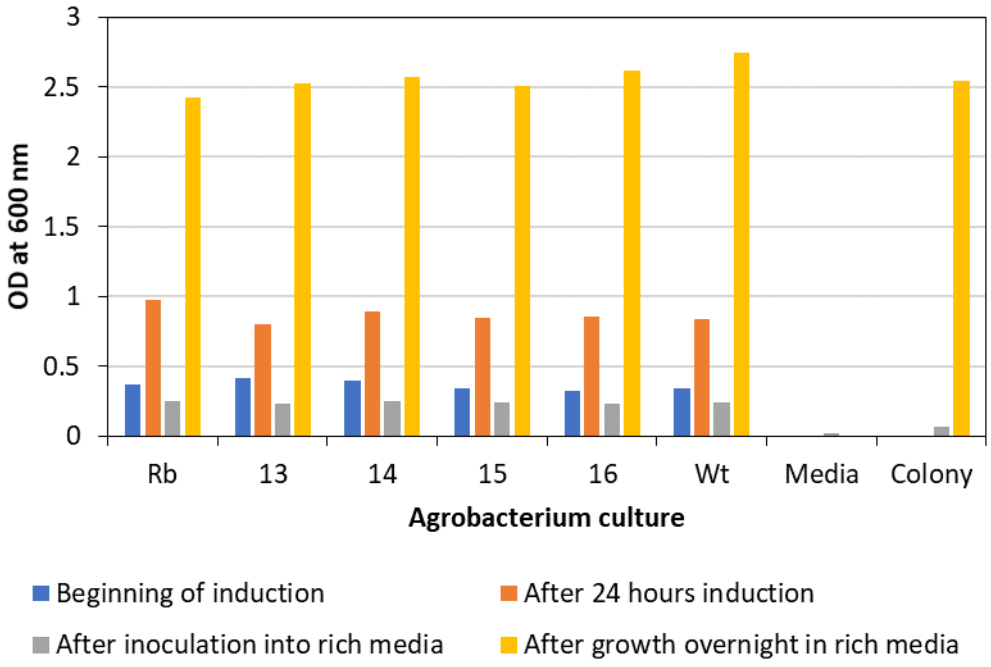


Figure 4-13 Viability of *Agrobacterium* strains after growth in induction medium. Wt: wild-type *Agrobacterium* cultured without antibiotic selection. Media: Un-inoculated growth media. 600 µl of induced culture was added to 6 ml rich medium + plasmid antibiotics to test cell viability. As a positive control, a fresh single colony from a plate of strain 13 was inoculated in to the same medium.

Since stressful conditions can apparently cause *Agrobacterium* sp. to lose its disarmed helper Ti plasmid (but not the binary vector, which is under antibiotic selection) which contains the virulence genes for transfection of the T-DNA, PCR was used to confirm the presence of this plasmid. In *Agrobacterium tumefaciens* strain EHA150 there is no antibiotic marker for the helper plasmid pTiBo as it is generally large enough to be stable. The chosen primers have been used in the literature to confirm the presence of pTiBo by giving a band of 766 bp (Deeba et al., 2014), and PCR carried out here resulted in bands with variable strengths at the expected size with clones "Rb" and "14" giving the strongest band (**Figure 4-14**). It is unclear whether the variable band strength is due to the quality of the template DNA, since another PCR reaction (gel not shown) was attempted using primers for the binary vector plasmid and gave no bands despite the cells growing in the presence of the antibiotic on the binary vector suggesting it must be being maintained in the cell. Further tests on the induced cells would be required to ensure the Ti plasmid is present, though the successful use of this induction protocol in the literature to generate microalgal transformants (Mini et al., 2018) suggests that it should be.

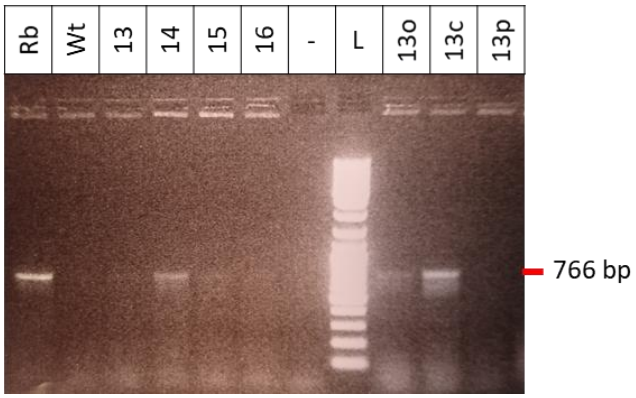


Figure 4-14 PCR to confirm presence of the Ti-plasmid of cells from cultures of induced *Agrobacterium*. Ladder is Thermo GeneRuler DNA Ladder Mix. Rb, 13, 14,15 and 16 are *A. tumefaciens* transformed with different binary vectors. Wt is *A. tumefaciens* without a binary vector. The template for these reactions is a small amount of liquid culture from cultures grown in induction medium overnight and concentrated by 3.5 times. 13o template is a culture of rich media which was inoculated using the induced culture to test viability of the cells. 13c template is a culture inoculated directly with a colony into rich media. 13p template is a colony touched from a plate stock into the PCR tube.

4.6.3 The ratio of *Agrobacterium* to algal cells during co-incubation

Agrobacterium sp. must have direct contact with the algal cell wall to penetrate and deliver the T-DNA as discussed in the Introduction. Therefore, there must be an adequate number of cells to enable this contact, but not too many that the *A. tumefaciens* overgrows and inhibits algal viability. The two ways to alter this ratio are via the amount of *A. tumefaciens* added to the co-incubation plate and the amount of algal pre-incubation. The amount of algal pre-incubation has been discussed previously.

To vary the amount of *Agrobacterium* sp. used in co-incubation, 200 µl of three *A. tumefaciens* cultures were added to the co-incubation plate at an OD_{600} of one (the standard concentration), five or ten. The cells were adjusted to these concentrations by spinning and resuspending in the required amount of medium. The number of colonies obtained on the selection plates were zero, eight and zero respectively, indicating that the best *A. tumefaciens* OD was five. However, due to the low numbers involved, there is not enough data to determine whether the differences seen are significant. Since the increased amount of *A. tumefaciens* did not seem to inhibit the algal growth then an OD of five was used in future transformations because this condition resulted in the highest number of colonies.

4.6.4 A liquid selection step is not useful because it is ineffective at removing background

Although the enrichment of putative transformants using the liquid selection step was the only way colonies were obtained in the test transformations (**Table 4-3**), there were still a large number of background colonies and the die-off during propagation rounds suggests that many of these may not be true transformants and instead became acclimatised to G418 during selection steps. To test the efficacy of the liquid selection step, a negative control transformation using the same test protocol as section 4.3.1 was carried out using just wild-type *A. tumefaciens* cells during co-incubation, which do not contain any binary vector for transfer of antibiotic resistance.

The liquid selection (G418 50 µg/ml) was carried out for 3 or 5 days before plating on selective (G418 350 µg/ml) and non-selective media (**Figure 4-15**). At day 3, the culture was not visibly green, but on day 5 it was. Already this suggests that the G418 concentration in the liquid culture may not have been adequate, although this disagrees with the result shown in the 12-well test (**Figure 4-10**) where only 30 µg/ml G418 was needed to inhibit *C. sorokiniana* growth over 11 days, unless the volume of culture in this test was not scalable. As seen in **Figure 4-15**C, some cells are able to be revived from the 3-day old selective liquid culture when plated on solid agar containing no antibiotics, suggesting that although the selective liquid culture was not green after 3 days, there are cells within it that are not killed. This could be because the action of the antibiotic G418 is through protein synthesis inhibition so needs longer to take effect.

From the plating of the liquid cultures on solid media it can be seen that after liquid selection in G418 at 50 µg/ml for 3 days, no colonies are visible on the selective agar plate (**Figure 4-15**B). However, if the selective liquid culture is left to incubate for 5 days (when it became green), then a lawn of cells is visible on the selective plate **Figure 4-15**D. This highlights a key disadvantage of the liquid selection in that if one successful clone arises then it quickly overtakes the whole culture leading to large amounts of background colonies, probably originating from only a small number of truly different strains, on the solid selection plates and hence could hinder identification of true transformants. Consequently, liquid selection was not included in further transformations.

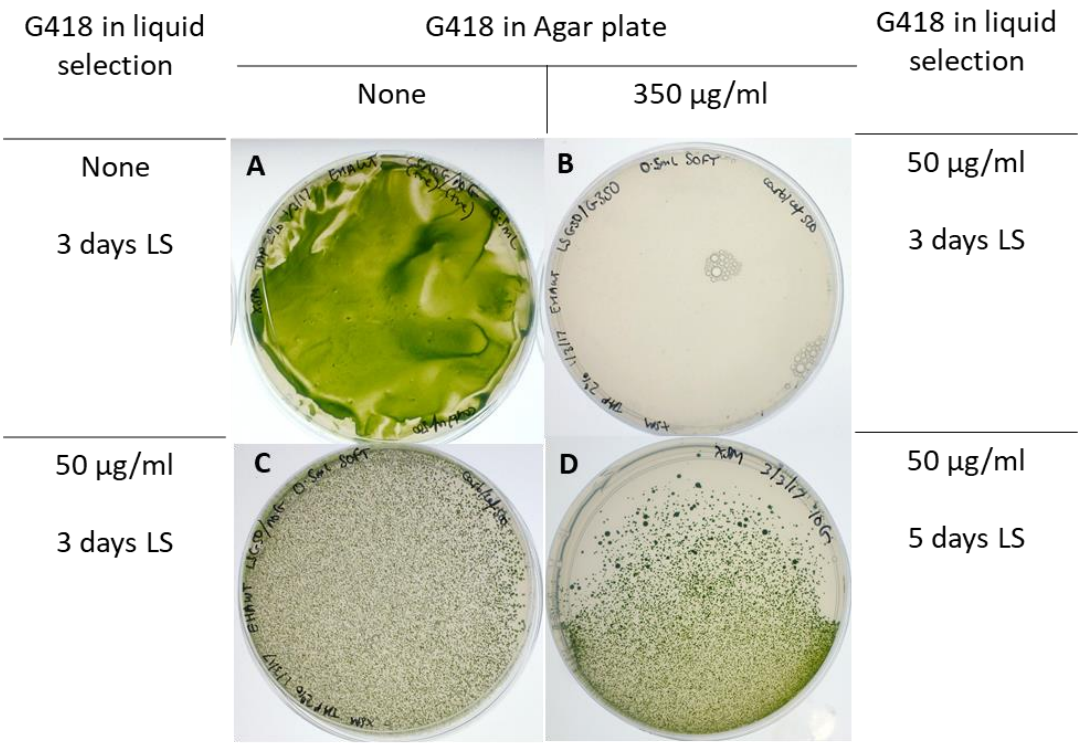


Figure 4-15 Survival of wild-type *Chlorella sorokiniana* UTEX 1230 after liquid selection with G418 antibiotic. Liquid cultures of recovered *C. sorokiniana* cells co-incubated with wild-type *A. tumefaciens* were incubated in the light at 25 °C for 3-5 days as specified. Part of the culture was plated on TAP agar containing the specified antibiotics. Plates were imaged after incubation in the light at 25 °C for 12 days (A, B and C) or 10 days (D).

4.6.5 Conclusion – efficiency remains low and needs to be increased for this to be a viable method for high-throughput applications

Two transformations were performed using the modified protocol with the control construct. The number of colonies and their survival is shown in **Table 4-4**. The efficiency is very low, and some transformants are lost by the second propagation.

Transfor-mation	Construct	Plate colonies (G350 selection)	Survival 1st round (G350)	Survival 2nd round (G350)
1	RbleC23 (+ve)	8	8	2 (1 strong)
	Wt Agro (-ve)	1	1 (weak)	0
2	RbleC23 (+ve)	0	-	-
	Wt Agro (-ve)	1	1	nd

Table 4-4 Number and survival of putative *Chlorella sorokiniana* transformants obtained using a modified *Agrobacterium transformation* protocol. G350 is G418 selection at 350 µg/ml. RbleC23 is the test construct and wild-type *Agrobacterium* contains the domesticated Ti plasmid but no Binary Vector.

A high efficiency transformation protocol is needed for genetic manipulation of microalgae to be successful. For example, it is often necessary to find a strain expressing the gene of interest to a high level, which due to the random integration of this transformation type may require screening of hundreds of colonies. Additionally, high colony numbers are needed if any test of promoter strength is desired in order to get a representative sample. Low transformation efficiency is a problem in microalgae, including for methods of *Agrobacterium*-mediated transformation. Coupled with loss of putative transformants over subsequent propagation cycles it makes it very difficult to make progress in genetic engineering. Additional tests involving molecular confirmation of each propagation round would be required to know whether these are loses of true transformants through silencing or whether they are elimination of false positives.

In this section, modifications to the initial protocol were explored. By reducing the number of algal cells at pre-incubation, and by removing the overnight recovery step, the amount of transformation culture plated out was doubled. There is still more room for improvement however and although in the antibiotic test section 350 µg/ml G418 was used it would perhaps be useful to test selection up to the maximum manufacturer recommendation of 500 µg/ml. Also, other selection markers could be considered since there are several as yet untested in *C. sorokiniana*. The addition of extra virulence steps and increasing the amount of *Agrobacterium* sp. in the co-culture were also modifications which looked promising. Further work would be needed to test other factors such as acetosyringone concentration, co-incubation

time, temperature and media. Although *Agrobacterium*-mediated transformation shows promise as a low-tech (though labour intensive) way of transforming microalgae, it is hindered by its low efficiency in this work, and although there are many reports in different algal species (see introduction), the method does not seem to be widely adopted over other methods such as biolistic, glass beads and electroporation.

4.7 Examination of transformant stability from pre-existing transformant lines

Stability of algal transformants has been an issue in algal transgenics with many studies reporting only transient expression (Bashir et al., 2016; Doron et al., 2016; Radakovits et al., 2010). *Agrobacterium*-mediated transformation is a nuclear transformation protocol that results in DNA integration into the genome, and therefore this genetic change is supposedly permanent (although may be silenced). Despite this, some studies have reported problems with instability of the selected phenotype. The reason behind this instability is not well characterised but it is reported that in the model *C. reinhardtii* that there can be silencing events through transcriptional repression by epigenetic mechanisms (Cerutti et al., 1997).

Previous work in the Purton lab (Hiegle, 2014) had generated putative *C. sorokiniana* transformants using *Agrobacterium*-mediated transformation and the RbleC23 construct (containing G418 as the selectable marker and Zeocin as a second reporter antibiotic, **Figure 4-1**). Further examination of these lines was carried out to assess the stability of *Agrobacterium*-mediated transformation. These tests were performed over 20 months after the putative transformants had first been isolated (Hiegle, 2014), and the lines had been maintained in low light on non-selective media after their initial selective propagations. There are six putative transformants and one escape colony which was isolated from a negative control plate during initial selection (Hiegle, 2014).

4.7.1 Spot tests show increased resistance to antibiotics than wild-type but so does an escape mutant

Spot tests were performed by incubating small scale cultures inoculated directly from an agar plate stock into non-selective liquid media for a day until mid-log phase (~OD 750 0.8-1). 10 µl spots of serial dilutions were then plated on TAP agar plates containing various G418 or Zeocin™ (**Figure 4-16**) and imaged after 7 days.

All strains were of similar dilutions as seen in the spots on the no antibiotic plate (**Figure 4-16**). Examination of the most diluted spots show between 1 and 13 colonies across the lines, corresponding to 1.2 - 16 × 10^4 cells per 10 µl spot or 1.2 – 16 × 10^6 cells/ml plating density. Therefore, the spot which compares best to the selection plating density of 100 million cells in 200 µl (corresponding to 5 million cells in 10 µl) is the third spot which should contain ~3-40 × 10^6 cells.

Examination of the spots on the G418 plates shows that despite all the strains growing better than wild-type at the low G418 concentration of 100 µg/ml, only transformant number 1 can grow at 350 µg/ml and above at all dilutions. Two other transformants 3 and 5 grow weakly in the most concentrated spot, and number 7 has a few surviving colonies, but this is similar to the negative escape line TP(-). Since *Agrobacterium* transformation is random integration it could be that the T-DNA in transformant number 1 has integrated into a more highly expressed region of the nuclear genome than transformants 3 and 5. Interestingly, this performance on G418 for transformant 1 is not matched by its ability to grow on zeocin, which is only as strong as the escape mutant TP(-) (**Figure 4-16**). This could be related to the construct only being partially present because of shearing, a problem reported in *Agrobacterium*-mediated transformation of *C. reinhardtii* (Mini et al., 2018). PCR would be required to test this.

All other transformants appear to grow better than the TP(-) escape mutant on zeocin despite their poorer performance on G418. Although it seems the markers may be expressed, the performance of all the strains has decreased compared to earlier tests when they were first isolated (Hiegle, 2014). Despite this all strains grow better than the wild-type in the antibiotic conditions suggesting they are true

transformants and the construct has not been lost from the genome. However, the reduced phenotype for all lines suggests there may be some silencing over time.

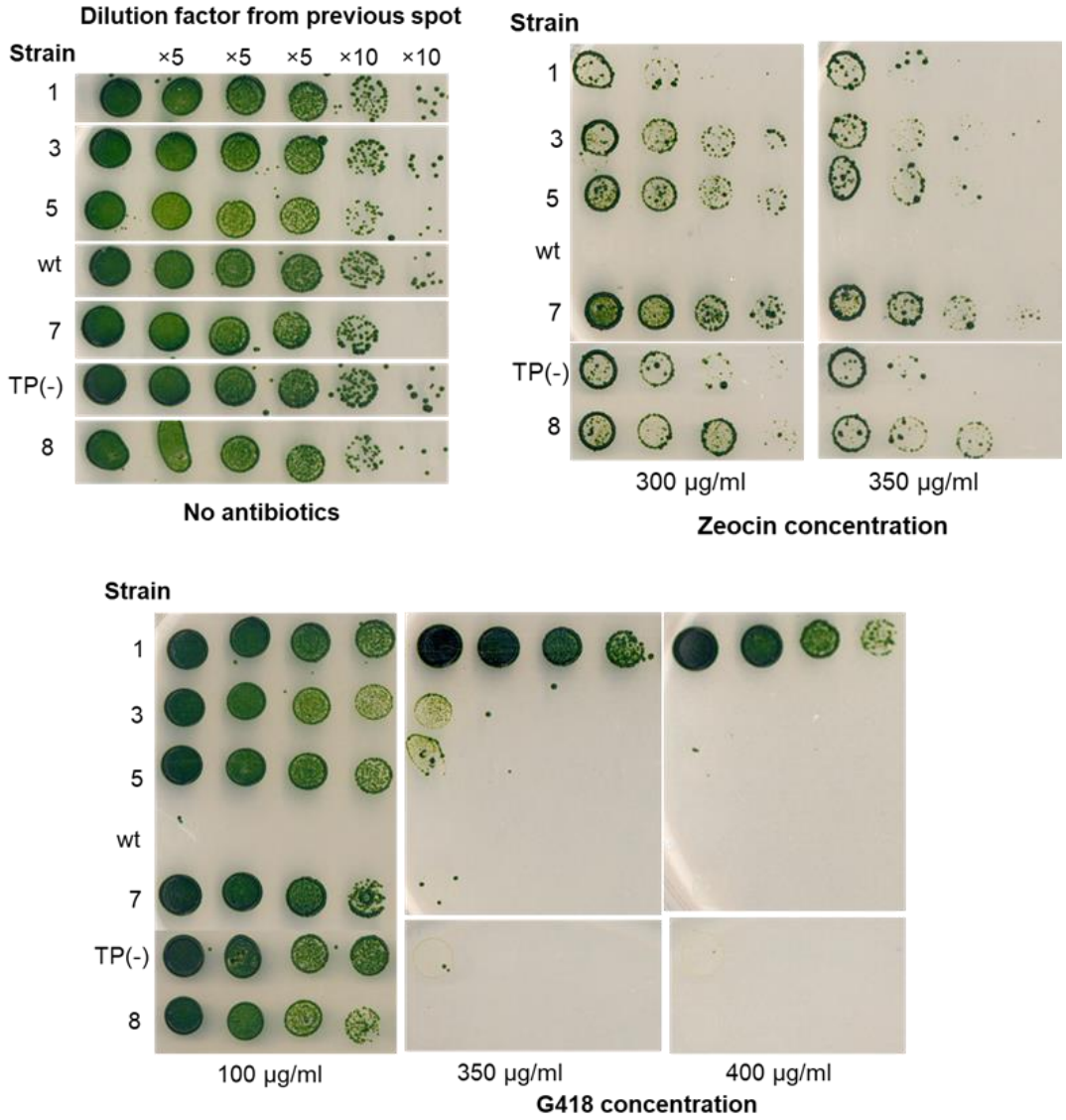


Figure 4-16 Spot tests on 20-month-old putative transformants of *Chlorella sorokiniana* containing G418 and zeocin resistance markers. Cultures were grown in 3 mL non-selective liquid media inoculated from an agar plate overnight in the light until mid-log phase. Serial 5 times dilutions were made (unless otherwise specified) and plated out in 10 µl spots on TAP agar containing different antibiotic concentrations and incubated in the light for 7 days before imaging. The putative transformant lines are strains 1,3,5,7 and 8. TP(-) is an escape colony isolated on a negative control plate during selection. wt is wild-type *Chlorella sorokiniana* UTEX 1230.

4.7.2 Southern blot confirms the presence of transgenic DNA in one of the old transformants

Molecular confirmation of transgene integration was attempted by Southern blot in collaboration with Noreen Hiegle. The probe was designed to bind to the terminator region of the *ble2I* cassette, generated by PCR (**Figure 4-17**) then gel extraction was performed and the final DNA concentration measured as 46 ng/µl.

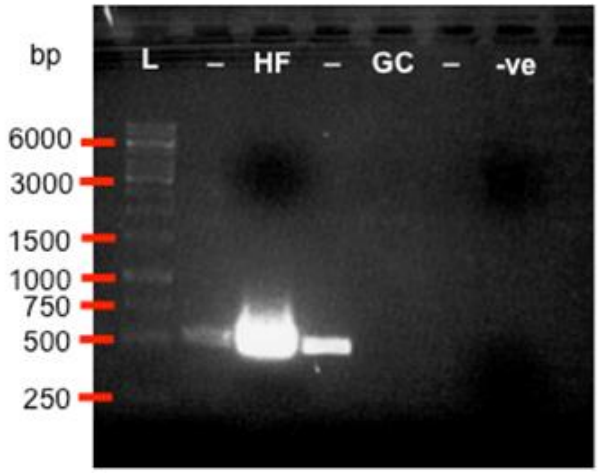


Figure 4-17. PCR to generate the DNA probe for Southern blot analysis. Expected product size 532 bp. Two different PCR buffers were used (HF or GC) from NEB. -ve contains no template. L is ladder GeneRuler 1kb (Thermo).

The gel-extracted DNA fragment was DIG-labelled using a commercial kit as described in section 2.2.8.2, and a spot test (**Figure 4-18**) confirmed the probe was adequately labelled compared to the positive control supplied with the kit.

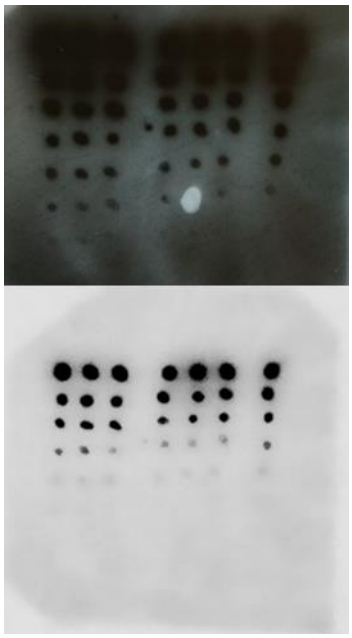


Figure 4-18 Testing labelling of the DNA probe for Southern blot. The designed DNA probe (first 3 lanes and last lane) labelling reaction is adequate since it is comparable to the control DNA supplied in the kit (middle three lanes). DNA concentration of spots descending is 10, 3, 1, 0.3, 0.1, 0.03, 0.01 and 0 pg/µl. Top: imaged with photographic film. Bottom: imaged with the Odyssey imager.

The Southern blot (in collaboration with Noreen Hiegle) was performed on genomic DNA extracted from three of the transformants described above: 3, 5, 7 and the wild-type as the negative control. Genomic DNA was digested with either XhoI or NcoI before blotting onto the membrane and probed using a DIG-HighPrime Kit. After probing and imaging, a single band was obtained for transformant 5 (**Figure 4-19**), but only when the digestion enzyme used was XhoI.

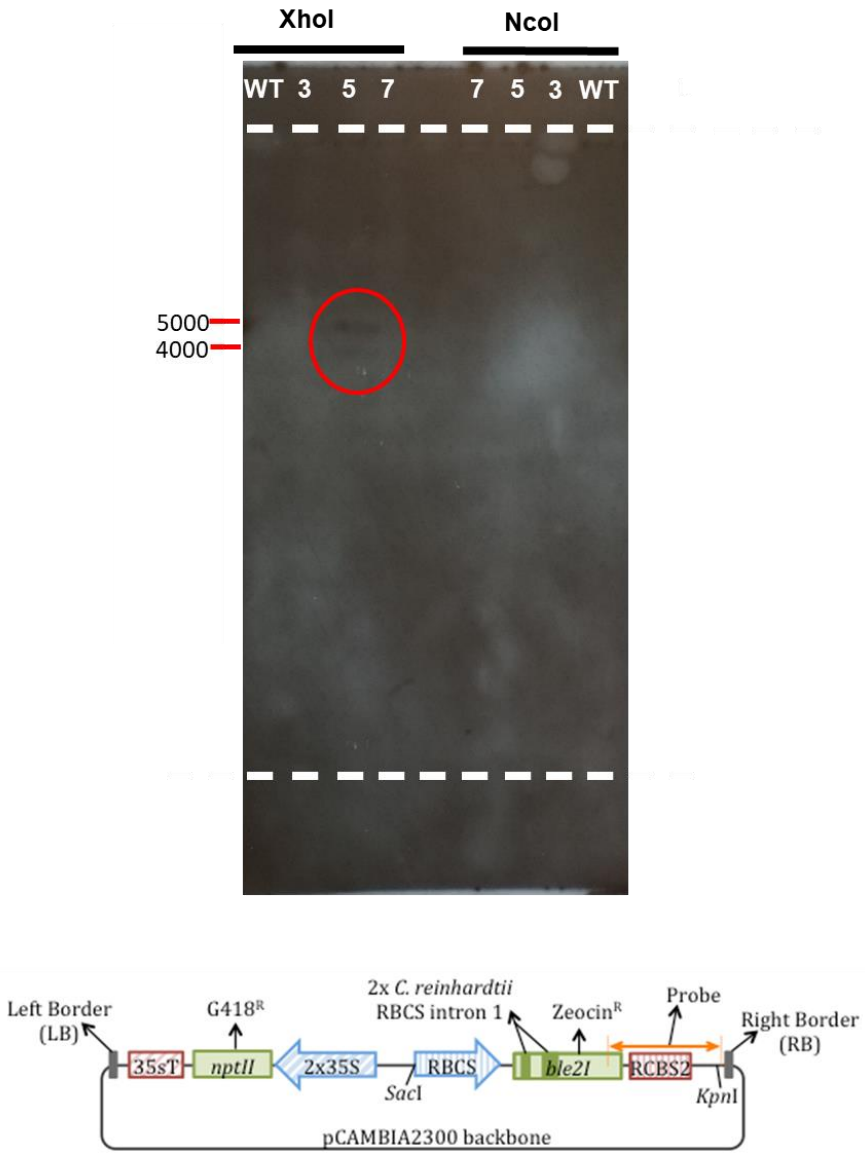


Figure 4-19 Southern blot of 20-month-old putative *C. sorokiniana* transformants. *XhoI* digested genomic DNA from 20-month-old putative transformants of *C. sorokiniana* shows a single band in transformant number 5 lane. The plasmid diagram is adapted from Hiegle 2014. The southern blot was carried out in collaboration with Noreen Hiegle.

Interestingly, number 5 was not the best performing transformant seen in spot tests (**Figure 4-16**). Southern blots of the other transformants were not performed due to time constrains but would be useful in future. In particular, it would be interesting to know whether the elevated performance in spot tests of putative transformant number 1 is due to multiple integrations or a single integration that has gone into a particularly high expressing region of the genome, since *Agrobacterium*-mediated transformation is random integration. The signal on the blot is very weak

and the blot would need to be repeated to confirm the absence of DNA integration in lines 3 and 7, especially given their good performance in spot tests. These lines have previously tested positive for transgene integration by PCR (Hiegle, 2014) so the absence of a Southern blot product for some is unexpected.

4.8 Discussion

4.8.1 The importance of a reliable selectable marker

The selection of putative transformants is a crucial part of the process of any transformation. Antibiotic resistance is the standard selection in lab work for the majority of microbial work and eukaryotic cell culture. *C. sorokiniana* has a particularly high resistance to commonly used antibiotics. Initial transformations were hindered by false-positives in which putative transformants were found to be selection survivors upon subsequent propagation rounds. Although this is a common problem in algal transformations and could also be due to transgene silencing, it is more likely due to many false transformants being picked due to the high background observed. This hinders the method and therefore experiments were undertaken to optimise the concentration and explore some of the potential sources of variation in effect of the antibiotic.

4.8.1.1 G418 and zeocin

In the experiments above, the concentrations required for selection were optimised and found to be sensitive to cell number in a non-linear fashion. It also led to modification of the transformation method to reduce the number of cells for selection and try to increase efficiency. Soft agar was found to be necessary to avoid 'hotspots' of resistance. Although G418 was only tested up to 400 µg/ml, it would have perhaps been useful to test higher concentrations. In the literature, one strain of *C. sorokiniana* was selected on 1 mg/ml (Hawkins and Nakamura, 1999) but they did not specify the number of cells.

The consistency of the selection and background varied between experiments. In addition to the factors tested, some of this could have been due to

storage of the antibiotics. Some antibiotics are very sensitive to storage conditions (Nickolai et al., 1985) and must be stored in single use aliquots and this was not done at the beginning of the PhD. G418 and Zeocin are reportedly subject to batch to batch variability and are sensitive to freeze-thaw (manufacturer advice). Some of the results on high concentrations from the test transformations may have been due to aliquots being used multiple times so the antibiotic was less effective.

It would have been useful to test the putative transformants after several rounds of propagation on non-selective media to see how stable the transformation was. Due to time constraints this was not carried out but could be useful in the future.

4.8.1.2 Other potential selectable markers

Although antibiotic selection is common there are disadvantages. In cloning, the resistance gene occupies space in the DNA vector. Also, for industrial applications, antibiotic resistance is undesirable and often there is a requirement to have a way of either removing the marker or selecting without the use of antibiotics. There are some interesting alternative markers which have been described, such as self-excising or heat-shock inducible markers (Dalla Costa et al., 2015), and increasing efforts towards so-called 'cisgenesis' and 'intragenesis' techniques (Galarza et al., 2016). Whilst recessive mutants are less suitable for non-characterised organisms as they require knockout strains for restoration of function, dominant mutants such as those showing resistance to a particular herbicide could be exploited by using the mutated gene to transform wildtype cells to herbicide resistance. For example, a mutant of *C. sorokiniana* that is resistant norflurazon has been isolated in the Purton lab (unpublished data). Since the herbicide is known to target phytoene desaturase (PDS) then it is likely that the *PDS* gene carries the resistance mutation. PCR amplification of PDS from the mutant would therefore provide a suitable dominant marker.

4.8.2 Transformation methods have multiple parameters that need optimising for each strain

Transformation methods are not well established amongst microalgae apart from a few model organisms such as *C. reinhardtii* and *P. tricornutum*. The large diversity of microalgae also makes it difficult to translate reported protocols from one species to another and the effectiveness of the different protocols vary greatly between strains and need optimising.

In this work, electroporation was only briefly attempted but it is the quicker and easier method so would be worth pursuing further in the future. The *Agrobacterium* method does have its own advantages which must be considered such as the ability to transfer large portions of DNA at mostly single or double copies. These advantages and disadvantages were recently highlighted in a publication by Mini *et al*. (2018) who directly compared these two methods in *C. reinhardtii*. As mentioned previously, they were unable to replicate the original *Agrobacterium* protocol (Kumar et al., 2004; Pratheesh et al., 2014) and the modifications they made to the protocol were the basis of some of those attempted in this chapter. In their comparison, Mini *et al*. 2018 show that *Agrobacterium*-mediated transformation gives more false-positives than electroporation, which they postulate could be due to degradation of the antibiotic in longer co-incubations or the antibiotic being inactivated by *Agrobacterium*. The results presented in this chapter also show a large number of false positives as identified by their inability to survive in subsequent propagation rounds.

Mini *et al.* 2018 also observe less silencing in transformants in *Agrobacterium*-mediated transformation, but both electroporation and *Agrobacterium* methods resulted in the majority of selected transformants having rearrangements in the reporter gene. In this chapter potential silencing is observed through decline of phenotype both in old transformants and in new transformations. Rearrangements and molecular confirmation of the putative transformants was not carried out but would be a priority for future work. Mini *et al.* 2018 also find that *Agrobacterium*-mediated transformants had their transgene in transcribed regions which means that

despite its drawbacks the method could be useful for insertional mutagenesis projects (Mini et al., 2018).

Therefore, it is worth carefully considering the transformation method depending on the purpose of the work and how many transformants are needed. Also, this chapter has only considered nuclear transformation. There have been a few reports of chloroplast transformation in microalgae outside the established model alga, *C. reinhardtii* although currently no reports for *Chlorella* sp. (Esland et al., 2018). This could be worth pursuing in the future depending on the desired application.

4.8.2.1 *Agrobacterium* method optimisation

In this work, various modifications from the original *Agrobacterium*-mediated protocol were tested. Relating the appropriate selection conditions to the transformation methodology was addressed through modifying the number of cells available to plate at the end of the transformation. This involved the removal of an overnight recovery step which was deemed unnecessary. Generally, recovery steps are used more in physical methods like glass beads, biolistics and electroporation, but they also related to the mode of the antibiotic to give the cells time to express the resistance marker. In the case of G418, since it is a protein synthesis inhibitor rather than causing active damage then recovery is less essential. Additionally, *Agrobacterium*-mediated transformation is presumably less physically stressful to the cell than the other physical methods, though this could be tested by imaging the bacterial-algal interaction via Scanning Electron Microscopy (SEM).

The consideration of proper *A. tumefaciens* virulence, which is neglected in most microalgal *A. tumefaciens* protocols, was modified. Although the cells were confirmed to be viable after this, PCR confirmation was unable to be obtained for presence of the helper plasmid, which does not have a selectable marker but is stable (being large and low copy) unless the cells are stressed. It is unclear whether this was a problem with the PCR reaction or whether the cultures had lost this plasmid. In which case, it would explain why there were no putative transformants for some transformations.

CHAPTER 5 MINING A CHLOROVIRUS GENOME FOR GENETIC PARTS FOR DRIVING NUCLEAR GENE EXPRESSION IN *CHLORELLA SOROKINIANA*

5.1 Introduction

The field of microalgae is still in its relative infancy in terms of genetic engineering compared to other organisms where these tools are well established. For microalgae, toolkits have been established but most efforts are focused on the model organism *Chlamydomonas reinhardtii*. There have however been increasing movements towards other industrial species of interest such as the diatom *P. tricornutum* and the eustigmatophytes *N. gaditana* and *N. oceanica*. One difficulty with microalgal genetic engineering is that due to the large diversity of microalgae across different phyla and classes, even between the models above, any part for one organism is not transferrable or guaranteed to behave the same in another microalga. Therefore, many efforts are sporadic and parts must be tested in each strain of interest. Even within the *Chlorella* genus there are problems because of the loose definition of *Chlorella* sp. (see introduction section 1.5.1) of which many are being reclassified.

There is an acknowledged lack of genetic parts for microalgae (Spicer and Purton, 2016) which limits progress in creating stable transgenic lines and fine control of transgenes through, for example, inducible systems. Homologous elements such as promoters can sometimes be preferable as they have been shown to perform better than heterologous elements in some reports (Hallmann, 2007). However, for microalgae which do not have a sequenced genome, this is not a viable strategy. One way this has been addressed is promoter trapping where a promoter-less antibiotic gene is transformed into the organism and then genomic context surrounding the gene for successfully growing clones is determined by PCR. This has been conducted in *C. reinhardtii* (Vila et al., 2012) and several other industrial species including *P. tricornutum*, *C. vulgaris*, *C. sorokiniana* and *Nannochloropsis* sp. (Spicer and Purton,

2016). However, a disadvantage of promoter trapping is the difficult process of characterising the 5' marker integration site by 5' RACE. Additionally, there have been efforts to create synthetic promoters based on known consensus core elements (Juven-Gershon et al., 2006). Although these synthetic promoters contain mostly conserved elements, there is some variation and hence synthetic promoter design can be tailored based on the motifs of a specific species. For example, the first synthetic core promoter specifically designed using the yeast *Pichia pastoris* was recently characterised (Vogl et al., 2014). Synthetic promoters have also been successfully employed for plants (Liu and Stewart, 2016) and the model microalga *C. reinhardtii* (Scranton et al., 2016) so could be a promising strategy for other microalgal species where sources of endogenous elements are not available. For *Chlorella* sp., there have been many individual reports on transgene expression using various promoters and terminators, but there has not been a collective effort as seen in the other species and also only a few of these reports relate to *C. sorokiniana* (as summarised in recents reviews (Liu and Chen, 2016; Yang et al., 2016)).

Linked to the development of a toolbox of parts is the need for a strong validation system and the characterisation of the behaviour of these DNA parts in different hosts (Scaife and Smith, 2016). This is often not conducted in a suitably rigorous manner to allow parts to be properly compared. However, this is not just a problem with microalgal parts, but across all organisms where a synthetic biology approach is to be taken. Observations may be specific to a particular system and precise characterisation is hindered by the intrinsic complexity of biological systems and their ability to adapt their metabolism to changes. Ways to combat this variability include high numbers of replicates and using internal standards within the cell or characterisation to a standard promotor for comparison across organisms. For example, this method is commonly used in *E. coli* (Beal et al., 2016).

The previous chapter examined a suitable transformation strategy for *C. sorokiniana* and highlighted significant challenges in this area. This chapter is concerned with the design of a library of DNA parts and considers how they would be tested. Preliminary evidence for their validation is presented, although much more work is needed in this area. It therefore provides a fertile ground for further research.

5.1.1 Types of genetic part – Levels of cellular control

In order to successfully engineer an organism, a robust and reliable set of genetic parts is needed. These genetic parts are generally derived from the existing eukaryotic gene structure and a complete expression cassette must take into account these different elements and how they impact upon the expression of a gene by the host. Gene expression is controlled at several stages, each determined by different parts of the DNA sequence (Alberts et al., 2008) (**Table 5-1**). Key parts include promoters, terminators, 5' and 3' untranslated regions (UTRs), intron/exon signals and the specific DNA sequence of the protein coding region (**Table 5-1**).

Genetic Element/Part	**Affected process in gene expression**
Promoter	Transcription of DNA to RNA via recruitment of transcription factors, activators and RNA polymerase. The 'core' promoter is generally considered to be 50-100 bp including the transcription start site
5' UTR	Translation of RNA into protein via recruitment of 5' mRNA cap and ribosome, recruitment of translation initiation factors.
Protein coding DNA sequence (CDS)	Translation rate, depending on codon usage of host and availability of tRNAs. mRNA structure can influence stability and translation.
Introns/Exons	Pre-mRNA processing: splicing
3' UTR	Pre-mRNA processing and translation; contains Poly-A signal for addition of tail which affects stability of the mRNA
Terminator	Transcription of DNA to RNA: signals 3' cleavage of transcript and Poly-A tail addition

Table 5-1 Key genetic parts involved in eukaryotic gene expression.

These parts all come together to provide a multitude of levels of control and feedback mechanisms for the cell. Working from the bottom up, firstly there is the availability of the chromosomal DNA by its packaging via histones and nucleosomes. Then there is the transcriptional control by the promoter, which may contain *cis* or *trans* acting factors. An example of this is the *C. reinhardtii* HSP70 enhancer which acts in *cis* to enhance the activity of promoters such as that from the commonly used RBCS2 promoter (Schroda et al., 2002, 2000). Then there is mRNA itself which can have different stability based on its folding dynamics and any secondary structure elements such as hairpins. Then there is translational control by ribosome

recruitment dynamics. Finally, the proteins themselves have different stabilities and degradation pathways related to their structure and function. In much of genetic engineering, the primary focus is on controlling transcription and using a strong promoter with a suitable terminator. Although transcription is only one part of the central dogma discussed above, it is the focus of the DNA parts in this chapter.

5.1.2 A synthetic biology approach to algal transgenics

One key paradigm in synthetic biology is the standardisation of genetic parts for efficient exchange and control. In order to align with this in plants, the different types of genetic element have been further split into 12 defined categories with specific flanking regions in a standardised plant syntax involving Golden Gate cloning (Patron et al., 2015). All of the genetic parts need to be correctly tuned for optimal expression of coding sequences, though of these the promoter and codon optimisation have a particularly high influence. Effects of codon optimisation have been demonstrated in microalgae (Heitzer et al., 2007). The use of intron sequences is a known way to enhance protein expression by affecting transcription (Baier et al., 2018; Hernandez-Garcia and Finer, 2014; Specht et al., 2010), though these should be homologous (or potentially synthetic (Hernandez-Garcia and Finer, 2014)) introns because heterologous ones may be spliced incorrectly (Hallmann, 2007).

In this chapter the elements of synthetic biology are considered by conforming to a common plant and algal cloning syntax (Crozet et al., 2018; Patron et al., 2015) and choosing appropriate reporters to allow relative characterisation of promoter strength.

5.1.3 Sources of genetic parts

Genetic parts can be derived from a variety of sources. In plant biology, promoters derived from plant viruses are very commonly used with high success; additionally, synthetic and chimeric promoters are also receiving increasing attention (Liu et al., 2013; Saunders and Lomonossoff, 2013). In microalgae, endogenous homologous promoters from highly expressed genes tend to produce better results than promoters derived from plant viruses such as the CaMV 35S promoter, which

has only limited success in algal systems (Akbari et al., 2014; Díaz-Santos et al., 2013; Specht et al., 2010). Synthetic promoters have the potential to be short and strong (by presenting natural parts in a 'condensed' form), can be engineered as regulatory devices to give expression profiles which do not exist naturally, and can be more resistant to homology-dependant gene silencing (Hernandez-Garcia and Finer, 2014; Liu et al., 2013; Liu and Stewart, 2016). However, complicated computational methods and/or large experimental validation screens are required to identify sequences to create such minimal 'core' promoters and the subsequent sequences to append to these for tailored levels of expression, though tools for this are becoming increasingly available (Davis et al., 2012; Kumari and Ware, 2013; Liu and Stewart, 2016). Synthetic promoters for *C. reinhardtii* have recently been designed and shown to perform better than endogenous ones (Scranton et al., 2016). Viral promoters have the potential for high and normally constitutive expression, but work most effectively in organisms related to the natural host of the virus (Liu et al., 2013; Saunders and Lomonossoff, 2013). They also serve as the 'core' backbone for some synthetic promoter designs (Liu and Stewart, 2016).

For these reasons, and the overwhelming success of plant virus DNA in plant biotechnology (Mushegian and Shepherd, 1995), algal viruses may be a useful source for novel DNA parts in microalgae molecular biology. Indeed, recently a putative promoter region from a diatom-infecting virus has been characterised and shown to give significantly higher transcription and translation than endogenous diatom promoters in *P. tricornutum* (Kadono et al., 2015). There is particular potential for *Chlorella* sp. in this regard since there is significant work in characterising *Chlorella* sp. viruses at the genomic and transcriptomic level, and there are well characterised techniques for identifying such *cis*-acting elements from genomic and transcriptomic data (Hernandez-Garcia and Finer, 2014). Several *Chlorella* sp. virus promoter regions have been characterised already in other hosts, as discussed in the section below.

5.1.4 Aims of this chapter

As discussed in the introduction to this Chapter, there is a need for well characterised DNA parts that are transferable between research groups to bring

microalgal genetic engineering up to the synthetic biology era. This chapter argues for the potential of mining chlorovirus genomes for genetic parts to use for heterologous gene expression in *Chlorella* sp., and uses published transcriptomic data to select two appropriate genes. Subsequent analysis of the upstream and downstream sequence to determine the promoter and terminator regions allows design of suitable parts.

5.2 Mining the PBCV-1 chlorovirus genome for novel promoters for use in *Chlorella*

There is a need for more novel DNA parts in microalgal biotechnology, especially for *Chlorella* sp.. By analysing published genomic and transcriptomic data of *Chlorella* sp. viruses we can identify heterologous regulatory sequences which could be used to drive gene expression in the algal host, as has been done in plants with plant virus parts.

5.2.1 Introduction to *Chlorella* viruses

5.2.1.1 Natural History and biology

Viruses of the genus *Chlorovirus*, also known as *Chlorella* viruses, are part of the family *Phycodnaviridae*: large ('giant', 160-560kb, with up to 600 coding sequences) linear dsDNA viruses of eukaryotic algae (Van Etten and Dunigan, 2012). Chloroviruses infect symbiotic *Chlorella*-like green algae ('zoochlorellae') that are found within the cells of *Paramecium bursaria* (a protozoan), *Hydra viridis* (a coelenterate) or *Acanthocystis turfacea* (a heliozoon) (Van Etten and Dunigan, 2012) (see figure **Figure 1-2** for image of *Chlorella* sp. with *P. bursaria*). Chloroviruses are plaque-forming, though since the host confers resistance to the virus, plaque assays are only possible on those zoochlorellae which can proliferate independently of the host (Van Etten and Dunigan, 2012). Chloroviruses exist ubiquitously in fresh-waters at high titres around the world and can be easily isolated from natural sources by microfiltration of fresh-water samples and applying them to virus-susceptible *Chlorella* sp. strains (Bubeck and Pfitzner, 2005; Nebraska Center for Virology and Van Etten Lab, 2012; Van Etten et al., 1985b, 1985a; Zhang et al., 1988). The first

chlorovirus was isolated in 1981 when the symbiotic zoochlorella of the hydrazoan *Hydra viridis* failed to grow outside the host because of infection from the virus, which caused lysis of the entire population in 12-20 hours (Meints et al., 1981). Subsequently, the now prototype species of the genus, PBCV-1 (*Paramecium bursaria Chlorella* virus 1) was first described in 1982 (Van Etten et al., 1982).

Studies on the chlorovirus infection cycle using the prototype PBCV-1 have revealed that the virus infect in a bacteriophage like manner (Meints et al., 1981; Van Etten and Dunigan, 2012). Firstly, a spike structure makes contact with the cell wall (this first contact with the host cell receptor confers the specificity of the virus to its host) and the cell wall of the host is degraded. The virus lipid membrane subsequently fuses with the host cell membrane which undergoes depolarisation, and the linear dsDNA genome and associated proteins are ejected into the host cell where they move to the nucleus for early transcription using host machinery (the virus does not encode its own RNA polymerase) (Agarkova et al., 2014; Blanc et al., 2014; Van Etten and Dunigan, 2012). With significant viral transcripts for 50 genes occurring as soon as seven minutes post infection (p.i.) (Blanc et al., 2014), this migration of DNA to the nucleus is not a trivial feat considering the dense complexity of the cytoskeleton of the eukaryotic cytoplasm which would limit diffusion by Brownian motion (Luby-Phelps, 2013), and suggests some components facilitating active transport of viral DNA to the nucleus (Blanc et al., 2014). This is also highlighted in discussions of the infection mechanism of *A. tumefaciens* and translocation of the T-complex in the host, where assistance of intracellular host transport machinery is thought to be very likely (Tzfira and Citovsky, 2006).

5.2.1.2 Potential implications in human health

There are now many *Chlorella* sp. virus strains that have been isolated. Interestingly, one (*Acanthocystis turfacea Chlorella* virus 1 (ACTV-1) (Bubeck and Pfitzner, 2005)) was recently implicated in human health: ACTV-1 DNA was present in throat swabs of humans and was associated with decreased cognitive performance (Yolken et al., 2014). This was also confirmed in mouse models, resulting in altered gene expression profiles in the hippocampus (Yolken et al., 2014), and physiological

changes have been detected in murine cells lines (Petro et al., 2015). The authors are confident that their results are not explainable by experimental issues like sample contamination (Kjartansdóttir et al., 2015; Yolken et al., 2015). This highlights that viruses in the environment thought not to infect humans may have biological effects (Yolken et al., 2014). This is surprising considering the apparently strict specific host range of the *Chlorella* sp. viruses (Bubeck and Pfitzner, 2005; Reisser et al., 1988; Van Etten and Dunigan, 2012; Yamada et al., 2006), though the precise mechanism for this specificity (apart from its association with viral attachment to cell wall) is unclear (Jeanniard et al., 2013). The implication of ACTV-1 in human health is perhaps particularly concerning for *Chlorella* sp. since its wide sale of biomass as food supplements in the health market. It poses the question as to whether the species cultivated are free from virus, and if they are not whether this poses any risk. Especially considering the relative stability of ACTV-1 virus particles being quite high (Bubeck and Pfitzner, 2005). In fact there has been a case report of psychosis thought to be related to the individual's consumption of 'sun-chlorella' (Selvaraj et al., 2013), though the species was quoted as *C. pyrenoidosa*, which is free-living and not naturally infected by *Chlorella* sp. viruses (Fan et al., 2015). It is worth being aware of these cases since they highlight that biological systems often have unexpected effects. It also emphasises the need to consider potential issues of the sources of genetic parts and any immunogenetic responses they may trigger.

5.2.1.3 *Chlorella* virus genomes and insights

The first *Chlorella*-virus genome sequence, for PBCV-1, was completed in 1997 (Li et al., 1997), but corrected and updated in 2012 (Dunigan et al., 2012). The PBCV-1 genome is 331 kb, has 416 predicted protein coding sequences and 11 tRNAs (Dunigan et al., 2012). Currently, there are a total of 41 chlorovirus genomes sequenced with genomes of 287–348 kb and 319–381 predicted coding genes, and which show a large degree of diversification (Jeanniard et al., 2013). *Chlorella* sp. virus strains are grouped according to their zoochlorellae hosts: 'NC64A' viruses of *Chlorella variabilis* Shihira et R. Krauss (referred to previously as *Chlorella* NC64A (Proschold et al., 2011)), of which the prototype PBCV-1 belongs; 'SAG' viruses of *C. heliozoae* (referred to previously as *Chlorella* SAG 3.83) and 'Pbi' viruses of

Micratinium conductrix (referred to previously as *Chlorella* Pbi) (Proschold et al., 2011).

Studies on PBCV-1 and its host *C. variabilis* NC64A, the genome of which has also been sequenced (Blanc et al., 2010), are an important model for algae virus-host dynamics. Transcriptional studies on both the algal host and virus during the course of infection have found temporal regulation of virus transcripts (Blanc et al., 2014; Rowe et al., 2014). The studies have also yielded insights into the ancient and long evolutionary history of the chloroviruses potentially back to division of prokaryotic and eukaryotic lineages (Kang et al., 2004; Yamada et al., 2006). Chloroviruses have even been implicated in influencing the evolution of their hosts, for example by horizontal gene transfer of genes (Blanc et al., 2010). Chloroviruses also possess many features unique from other viruses including encoding of interesting proteins, as discussed below.

5.2.1.4 Interesting genes and features of chlorovirus and their biotechnological uses

Chlorella sp. viruses have been noted as containing the smallest known, or among the smallest proteins of particular classes, and sometimes unique proteins not found in any other viruses. These include eukaryotic ATP-dependant DNA ligase (Odell et al., 2003), type II DNA topoisomerase, a histone methyltransferase (vSET (Mujtaba et al., 2008)), and a K^+ channel (Kcv (Kang et al., 2004; Romani et al., 2013)), reviewed in (Van Etten and Dunigan, 2012; Yamada et al., 2006). Of these the K^+ channel has received much interest as a minimal catalytic unit (Kang et al., 2004).

Other interesting gene products include among others: polyamine metabolism, chitinases and hyaluronan synthase (reviewed in (Van Etten and Dunigan, 2012; Yamada et al., 2006)). In particular, chloroviruses are known for encoding an unusually large suite of proteins for carbohydrate modification (Van Etten et al., 2017) and uniquely encode machinery to glycosylate their own proteins, and even do this in unique patterns (De Castro et al., 2015, 2013, Van Etten et al., 2017, 2010). PBCV-1 also encodes 11 tRNAs (Cho et al., 2002) which the rare virally-encoded ribonuclease III is suggested to be implicated in processing (Zhang et al., 2003).

Some of these interesting products have been useful for molecular biology studies in other organisms since they function recombinantly, such as the DNA ligase (Kubota et al., 2015; Odell et al., 2003). From a biotechnological point of view, the *Chlorella* sp. virus aquaglyceroporin (aqpv1) has been expressed in tobacco for drought resistance (Bihmidine et al., 2014), and the DNA ligase has been suggested as a more efficient alternative to other ligases for annealing ssDNA to RNA splints in oligonucleotide probing assays (Lohman et al., 2014). The use of an intact viral particle itself as an infection agent to induce lysis in *Chlorella* sp. for easier extraction of compounds like lipids has been explored (Cheng and Labavitch, 2015), as has the use of infected cells for the production of the valuable polysaccharide hyaluronan (Rakkhumkaew et al., 2013). Their large genomes also mean that if chloroviruses were be to used themselves as DNA delivery vectors, they could be more useful for such a purpose than plant viruses, which often have small genomes (Henry and Meints, 1994).

5.2.2 *Chlorella* virus promoters used in the literature

The use of promoters from chlorovirus is not a new idea in itself, though reports are sporadic. A popular choice of gene in previous studies has been the promoter and 5'/3'-UTR region of a 33 kDa protein encoded by gene *A312L* (**Table 5-2**). This gene is of unknown function; despite one study naming it as an aspartyl-tRNA synthetase (Kawasaki et al., 2004), all other studies (Blanc et al., 2014; Graves and Meints, 1992; Kawasaki et al., 2004; Yanai-Balser et al., 2010) and the current NCBI record note that it as a hypothetical unknown protein. *A312L* has the highest transcript levels of any *Chlorella* sp. virus gene in several studies (Blanc et al., 2014; Graves and Meints, 1992; Kawasaki et al., 2004; Yanai-Balser et al., 2010) and was popular because of this and its apparent categorisation as a temporally early expression which is preferred because they are more likely to be able to be recognised by host transcriptional and translational machinery (Nguyen et al., 2009; Park and Choi, 2004; Park et al., 2005). It has the longest 5'UTR of any of the chlorovirus genes (Nguyen et al., 2009) and contains a 16x 5'-CAAA-3' motif (Kawasaki et al., 2004) (**Table 5-2**). When the promoter region and 3'UTR were used to drive protein expression in transformed *C. ellipsoidea*, protein levels were

comparable to those obtained using the CaMV 35S promoter (Park and Choi, 2004; Park et al., 2005). In *Arabidopsis* sp., the 5'UTR served as a translational enhancer by 15x when fused to CaMV 35S (Nguyen et al., 2009).

Another successful chlorovirus promoter actively used in research has been from the adenine-methyltransferase gene *amt* (**Table 5-2**). Using this promoter, higher expression of the gene of interest was reported in plants than with the CaMV35S promoter (Mitra et al., 1994; Mitra and Higgins, 1994). In this case, it is unclear why, since it does not contain a typical promoter motif (Mitra and Higgins, 1994) and is not among the most highly transcribed genes at any temporal category (Blanc et al., 2014). This promoter has also been tested in microalgae, with similar expression levels to the CaMV35S and RBCS2 promoters found in *Chlorella* sp. in one study (Hawkins and Nakamura, 1999) but enhanced expression in *Chlorella* sp. compared to CaMV35S in another (Kang et al., 2000). However, in *C. reinhardtii* when this *amt* promoter and the major capsid protein Vp54 promoter were used to drive gene expression (Ruecker et al., 2008), no expression was detected in the author's luciferase assay.

The success of these examples in driving gene expression, and in some cases enhancing gene expression than more commonly used elements such as CaMV35S, is evidence that mining the *Chlorella* sp. virus is a promising strategy for identifying highly active promoters.

5.2.3 Factors to consider in gene choice selectionr and strategy to use transcriptomic and motif data to determine the regulatory regions

Chlorovirus PBCV-1 has 416 predicted coding sequences (Dunigan et al., 2012). In order to select a putative promoter sequence for use in a heterologous organism one must decide which gene to take it from. The criteria used for this decision include high transcription levels and early temporal class in the infection cycle. The early temporal class is important because it means the regulatory sequences are more likely to be able to function using the host machinery and do not require a virally-encoded factor which would be unavailable when used in the heterologous system. Analysis in this way is made possible by several microarray and

transcriptomic studies including temporal motif analysis which have been published (Blanc et al., 2014; Fitzgerald et al., 2008; Yanai-Balser et al., 2010).

5.2.3.1 High transcription levels

Choosing a highly expressed or transcribed gene is the most obvious criterion for selecting a gene candidate for finding a new potential promoter. Although gene transcription is not necessarily directly related to protein yield because of post-transcriptional and post-translational regulation, it is still a commonly used strategy. Consideration of all the levels of control would require significant computational analysis and a depth of understanding of genetic control mechanisms in microalgae (Shrestha et al., 2013) which are not fully elucidated and is beyond the scope of this project. Therefore, the strategy used in this chapter focuses on the control level of transcription only.

It has been shown in a study of gene transcription over the course of infection for the chlorovirus PBCV-1 and its host *Chlorella variabilis* NC64A that the highest level of viral transcript produced is significantly more than the highest transcript level for host nuclear genes at 20 to 60 min p.i., by up to 37 times (Blanc et al., 2014). The high transcript levels could mean there are potentially some exciting 'super-promoters' in the chlorovirus genome, or that the transcripts are particularly stable and do not degrade as quickly, both of which would be ideal characteristics for use in heterologous organisms Also, some early transcripts accumulate to levels comparable to highly expressed transcripts in cellular organisms by t = 7 min p.i. (Blanc et al., 2014).

5.2.3.2 Temporal expression pattern

The temporal expression pattern should also be taken into account when selecting a viral promoter (Heoy-Kung et al., 2006; Park and Choi, 2004; Park et al., 2005). Studies on the dynamics of *Chlorella* sp. virus gene transcription show genes fall into categories of immediate early, early and early/late genes (Blanc et al., 2014; Kawasaki et al., 2004; Yanai-Balser et al., 2010). Early genes are preferred because at this stage they are transcribed by the host RNA polymerase and not by that

encoded by the virus. These early genes (such as transcription factors, mRNA processing proteins, proteins affecting translation and proteins involved polysaccharide synthesis (Kawasaki et al., 2004)) therefore rely on recognition by host machinery and so will more likely be recognised in heterologous hosts without any requirement of virally-encoded effectors encoded by late genes (Heoy-Kung et al., 2006; Park and Choi, 2004; Park et al., 2005).

Further evidence supporting early-gene choice is that most of the late genes seem to be dependent on initiation of viral DNA synthesis, because they were found to be blocked by the DNA replication inhibitor aphidicolin in one microarray transcript study (Yanai-Balser et al., 2010). Also, considering the fast speed at which viral effectors may come into play (PBCV-1 begins to halt gene expression, degrade host DNA and switch to viral transcription within 5 min p.i. (Agarkova et al., 2006)), the immediate-early genes will be the least likely to rely on viral processes. Common promoter elements of immediate early genes include a TATA-box and a 5'-ATGACAA-3' motif which may be repeated several times within -160 bp from the start codon (Kawasaki et al., 2004), or 5'-AATGACA-3' motif at -60 to -90 (Fitzgerald et al., 2008) (**Table 5-2**). Hence, these motifs are useful to consider in the selection of target genes as sources of promoters and the creation of synthetic promoters.

Gene/Motif	Notes	Reference
	Genes	
amt promoter	851bp upstream of translation start site was used. Performs better than CaMV 35S promoter in transgenic *Arabidopsis,* potato plants, maize, *Sorghum*, tobacco (15x higher) and transient wheat and rice (15-20x higher); also good function in prokaryotes *E. coli, Erwinia, Pseudomonas, Xanthomonas*. Interestingly, does not contain the typical TATA/CCAAT (eukaryotic) or -10/-35 (prokaryotic) motifs. In *Chlorella sorokiniana* and *Chlorella vulgaris* expression did not differ to CaMV 35s or *Chlamydomonas rbcs2* promoters. One study was unable to achieve any expression of luciferase in *Chlamydomonas.*	(Hawkins and Nakamura, 1999; Mitra et al., 1994; Mitra and Higgins, 1994; Ruecker et al., 2008)
DNA polymerase, DNA ligase, Chitinase	Chosen by authors because thought to be temporally early, though they do not appear particularly high in Blanc et al. 2014 study. Performed as well as CaMV 35S promoter in *Chlorella vulgaris* transformations.	(Heoy-Kung et al., 2006)
a312L (contains a 16x 5'-CAAA-3' motif repeat;	The most highly abundant transcript over the course of *Chlorella* virus infection. Encodes an aspartyl-tRNA synthetase. Promoter and 5'/3'UTR from this gene has comparable expression to CaMV 35S promoter in *Chlorella ellipsoidea*. Conflicting temporal classification (see text).	(Blanc et al., 2014; Graves and Meints, 1992; Kawasaki et al., 2004; Nguyen et al., 2009; Park and Choi, 2004; Park et al., 2005; Yanai-Balser et al., 2010)
	Motifs	
5'-AATGACA at -60 to -90;	Strongly associated with early gene expression. Resembles -35 element (TTGACA) in *E. coli* promoters. The authors identify 2 other promoter motifs (5'-ARNTTAANA-3' at -15 to -45; 5'-GTNGATAYR-3' at -50 to -80) but these do not have significant temporal patterns so are less useful.	(Fitzgerald et al., 2008)
5'-AAAAATAnTT	Present in Kcv promoter region of 16 *Chlorella* viruses	(Kang et al., 2004)
TATA-box (5'-TATAAAT) and 5'-ATGACAA	Typical promoter elements in 20 out of 23 immediate-early genes. 5'-ATGACAA can be repeated.	(Kawasaki et al., 2004)

Table 5-2 Notable genes and DNA motifs from chlorovirus PBCV-1. These promoters and 5'/3' UTR elements could be used as sources of, or in the design of, DNA parts for the genetic engineering of *Chlorella* sp. Kcv: functional K^+ channel; amt: adenine-methyltransferase.

5.2.4 Choice of chlorovirus genes for new expression elements

5.2.4.1 The genes A208R and A067R are selected

The success of the few attempts at using chlorovirus promoters in microalgae so far has been limited. The aim of this work is to identify a new more highly transcribed promoter for *Chlorella* sp. Considering the detailed transcription profiles from Blanc et al. 2014 (Blanc et al., 2014) (which can be used to show that the assumed 'early' status of gene promoters used in previous studies might not be accurate), identification of new candidate genes should consider the early temporal class as the most important factor. Using the raw data from Blanc et al. 2014, the top seven genes with the highest normalized transcript level were highlighted for each timepoint: in yellow for the top three genes, followed by orange and red (**Figure 5-1**). After ordering the annotated dataset by transcript level abundance at the first timepoint, the abundance patterns for the top genes shown in **Figure 5-1** were taken into consideration to select the candidate genes. Looking at the **Figure 5-1**, the top three genes were A088R, A208R and A131L but not all of these were chosen as discussed below. The two new candidate genes chosen were A208R and A067R, and do not appear have not been used or characterised to-date. These two were chosen based on their high transcription and presence as immediate or early genes after disregarding some other candidates as discussed below.

id	NORMALIZED MRPN T7'	NORMALIZED MRPN T14'	NORMALIZED MRPN T20'	NORMALIZED MRPN T40'	NORMALIZED MRPN T60'	Virion-associated (Dunigan et al. 2012)	Temporal class (Yanai-Balser et al. 2010)	Immediate early gene (Kawasaki et al., 2004)	AATGACA promoter motif	Expressed w/ aphidicolin (Yanai-Balser et al. 2010)	Predicted poly(A) cleavage site
A088R	1,241.3	3,822.5	4,904.6	2,864.5	1,517.1		Early	Yes	Yes	Yes	
A208R	1,114.9	4,917.8	7,971.2	9,482.5	4,882.4		Early	Yes	Yes	Yes	
A131L	948.3	3,398.5	3,785.3	1,741.0	804.6		Early			Yes	Yes
A341L	935.7	2,490.8	3,534.5	4,233.7	2,624.2		Early		Yes	Yes	
A067R	874.4	4,091.7	7,586.5	14,567.2	8,723.1		Early	Yes	Yes	Yes	Yes
A130aR	681.0	2,459.4	2,952.2	1,251.2	629.0						
A687R	660.0	2,454.4	3,939.5	5,342.0	4,677.2		Early			Yes	Yes
A161R	631.4	3,890.8	5,428.4	7,703.3	4,164.4		Early			Yes	Yes
A593R	590.2	2,760.4	4,517.3	4,503.5	3,712.1			Yes	Yes		
A548L	543.0	3,670.4	4,244.4	853.7	565.5	Yes	Early	Yes	Yes	Yes	Yes
A348R	500.8	2,853.3	5,148.4	8,454.1	4,142.1		Early		Yes	Yes	Yes
A005R	451.2	1,843.2	2,018.9	589.7	435.0		Early	Yes	Yes	Yes	
A084L	336.1	2,037.6	3,364.3	2,463.4	1,019.4		Early			Yes	
A130R	333.7	1,203.7	1,537.7	576.8	363.8		Early-Late			Yes	
A312L	301.4	2,128.7	6,839.4	51,563.9	165,053.0		Early-Late	Yes		Yes	Yes
A157L	280.8	1,522.1	915.5	700.1	482.7	Yes	Early-Late			No	Yes
A094L	276.2	1,778.0	3,562.8	4,577.1	2,202.7		Early		Yes	Yes	

Figure 5-1 Highly transcribed chlorovirus genes across the infection cycle. Table is taken directly from the supplementary data of Blanc *et al.* (2014) but modified to change the order and include colour coding indicating transcript levels in the top 3 (yellow), 5 (orange) or 7 (red). Time points in their study are at 7, 14, 20, 40 and 60 minutes post-infection.

A208R remains in the top four genes for transcript levels from T=7 to T=40 min p.i., and is in top 10 at T=60. A067R is in the top three for T=14-40, and in the top 5 for T=7 and T=60 (Blanc et al., 2014) (**Figure 5-1**). An earlier study also identified these genes as immediate-early (occurring at 5 min post infection and peaking at 40 minutes) (Kawasaki et al., 2004). Although the highest transcript level at the earliest T=7 time point is A088R, the chosen A208R and A067R transcript levels remained in the top five for longer, suggesting they are more stable across different cellular environments so may be more robust within a heterologous system. This was also why A131L and A314L were not chosen since they both reduced in level very quickly. By T=14 and T=20 both A208R and A607R are the highest transcript levels of all genes, and A208R is the only gene to remain in the top three transcript levels for the first three time points. Although A067R is only fifth highest transcript level at the first time point T=7, by T=14 it reaches the second highest transcript level and is the only one out of the (at least) top seventeen T=7 transcripts which remains in the top five by T=60. It could be useful to try A088R and A131L in the future, but this was not considered in this project due to time constraints.

The polyA tails of A208R and A067R were characterised by a previous study (Kawasaki et al., 2004) by examination of the mRNA transcript size over time, and were found to reduce after 10 or 20 minutes post-infection respectively. Whether this is a virally or host-controlled process is unknown.

It is also interesting to see gene A312L (which has been previously used in the literature (**Table 5-2**)) appears as the most highly transcribed gene of all across the whole infection cycle, presumably the reason it was used by those authors. However, it is evident from the table that it is not highly transcribed early in the infection cycle (despite conflicting evidence (Kawasaki et al., 2004)) indicating it might not be the ideal candidate. Several of the other most highly transcribed genes such as A441L, A440L and A439aR were not considered because their high levels were at T=60 and they had comparatively very poor levels at T=7 or T=14, which means they are likely relying on a virally encoded factor which would be unavailable when used in a heterologous system.

According to the deposited sequence in NCBI (Dunigan et al., 2012), A208R and A067R encode hypothetical proteins of 312 and 310 amino acids respectively. However one study (Kawasaki et al., 2004) states they have homology to human nucleolin (involved in RNA metabolism) and procyclin precursor (a coat protein of the parasite *Trypanosoma*), respectively. Neither of the protein products appear within the packaged virion itself, which is characterised by products of later-transcribed genes (Dunigan et al., 2012).

5.2.5 Bioinformatics analysis to determine promoter and terminator regions of the chosen genes

For the two chosen genes A208R and A067R, the next step was to decide which portion of the surrounding context corresponds to their regulatory sequences and examine them to check that they contain the expected control elements or DNA motifs associated with transcription regulation. In order to determine the actual portion of the PBCV-1 upstream and downstream A208R and A067R sequences to take as the putative promoter and terminator, the genomic context was considered and the sequence was analysed for the location of DNA motifs associated with core

promoter elements (**Table 5-1**) and transcription motifs at different temporal stages (Fitzgerald et al., 2008) (**Table 5-2**). Additionally, BLAST was used against the published transcriptomic data to obtain an approximation of the start and end of the transcripts

5.2.5.1 Genomic context

When determining the promoter or terminator sequence of a gene, one approach is to simply take a certain number of nucleotides upstream of the start codon and downstream of the stop codon. However, on examination of the upstream and downstream DNA context of the A208R and A607R genes it is of note that the genome is very tightly packed. For example, there is only 30 bases between the stop codon of the previous gene and the first putative start codon of A208R and 170 bases for A067R (**Figure 5-2** and **Figure 5-3**). Downstream of the genes, there is only 21 bases to the next gene for A208R and 30 bases for A067R. Because of this closeness, the selected sequence for the A208R promoter contains part of the upstream gene. For A067R though the chosen promoter sequence chosen does not overlap with the upstream gene because the motifs mainly occurred just in this intergenic region. The overlapping of genes in chlorovirus PBCV-1 was identified early-on in its characterisation, with studies using DNA probes (Schuster et al., 1990).

5.2.5.2 Using BLAST to determine the approximate transcript start and end

The NCBI Nucleotide BLAST tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Zhang et al., 2000) was used to determine the approximate transcription start and end sites. This was done by BLAST analysis of the the relevant portion of the PBCV-1 A208R and A067R genes against the transcript reads (Blanc et al., 2014) who had deposited their data in the Sequence Read Archives (SRA) in Bioproject PRJNA210187 as study SRP026413. The data from the T=20 time point was used that corresponds to blasting against experiment SRX317060 and the default parameters for highly similar sequences (megablast) were used except to change the maximum number of target sequences to 20,000. The data from 1000 reads was visualised in the NCBI graphic summary and the most 5’ and 3’ sequence was taken to be the approximate start or end of the transcript respectively (see **Figure A-1** and **Figure A-2** in the

Appendix). About 20 bases either side of the 5' end was used to annotate the approximate transcript start site as shown by the orange bars in context with other elements in **Figure 5-2**. Confusingly, for A208R the transcript seemed to start after the start codon suggesting that the annotated gene product may be incorrect or that it is in fact the shorter second putative ORF which is used. For A067R the transcript starts close to the ATG to include the putative Kozak sequence (discussed later) within a short 5'UTR of ~10-15 bases.

5.2.5.3 Analysis of promoter motifs

The promotor's function is to recruit the transcription machinery, and involves recognition of certain DNA motifs. Among these there are core promotor motifs that serve as binding sites for transcription initiator complexes, although not all motifs are necessary for successful transcription. The length of a core promoter is generally classified as the transcription start site (TSS) +/– 50 bases (Roy and Singer, 2015), but one plant promoter analysis extended this to +/– 500 bases to computationally explore elements across monocot and dicot plants (Kumari and Ware, 2013). In the chlorovirus transcriptional studies, the promoter region was defined as –150 to +50 bases around the AUG (Blanc et al., 2014; Fitzgerald et al., 2008). In addition to the temporal motifs specific to chloroviruses (**Table 5-2**), the region around the ATG was analysed for core promoter sequences including the initiator (Inr), downstream promoter element (DPE), TATA-box, CAAT-box and GATA-box (**Table 5-3**). The presence of the Kozak sequence, which is involved in translation initiation once the ribosome has bound to the 5'cap of the mRNA, was also considered. The purpose of identifying these motifs and annotating them in the chlorovirus DNA sequence (**Figure 5-2**, **Table 5-4**) was so that they could be included in the chosen sequences and aligned with algal coding regions when cloning.

Motif Name	Consensus sequence	Description	References
Transcription Initiation			
TATA-Box	TATAWAW	Typically -30 bp to TSS. Binds transcription factor TBP of RNApolII	(Alberts et al., 2008; Roy and Singer, 2015)
CAAT-box	nCAAT	Mean distance from TSS in plants 75 bp on both strands. Occurs in forward or reverse orientation.	(Müller et al., 1992; Roberto Mantovani, 1998; Shahmuradov et al., 2003)
GATA-box	WGATAR, HGATAR	Binding side of GATA family transcription factors (zinc-fingers). Present in eukaryotes including green algae but not Stramenopiles.	(Lowry and Atchley, 2000; Merika and Orkin, 1993; Thiriet-Rupert et al., 2016)
Inr (initiator)	YYANWYY Plants: WnWYMW	Typically spans the TSS itself. Binds transcription factors.	(Alberts et al., 2008; Shahmuradov et al., 2003; Smale and Baltimore, 1989)
DPE (downstream promoter element)	RGWCGTG	Located +28-32 of TSS. Binds TFIID. Requires the presence of Inr; does not act independently. Typically appears in TATA-less promoters.	(Alberts et al., 2008; Roy and Singer, 2015)
Translation			
Kozak sequence	Chlamydomonas: GCCRMCATGSCG Dicots: ..AAAAMAAUGGCU	Involved in translation initiation and can influence gene expression.	(Cross, 2016; Ikeda and Miyasaka, 1998; Joshi et al., 1997; Nakagawa et al., 2008)
Transcription Termination			
PolyA signal	AATAAA or TGTAA	Polyadenylation signal	(Blanc et al., 2014; Kawasaki et al., 2004; Wodniok et al., 2007; Zhao et al., 2014)

Table 5-3 DNA motifs in core promoter, translation initiation and transcription termination. The presence of these motifs are examined in the putative promoter and terminator regions for chlorovirus A208R and A067R. W: A or T; R: A or G; Y: C or T; H: A, C or T; M: A or C; S: G or C; N: any nucleotide. TSS: Transcription Start Site.

The motifs detailed below were annotated for visual representation in **Figure 5-2** (for A067R), **Figure 5-3** (for A208R), and in the detailed sequence information in **Table 5-4** (for both genes).

TATA box. Both A208R and A067R contain a TATA box (**Figure 5-2**, **Figure 5-3**, **Table 5-4**). In A067R the TATA box ends –49 bases from the start codon. For A208R the identified TATA-box actually begins +15 bases after the putative start codon, despite its typical location being –30 bases from a transcription start site (**Table 5-3**).

CAAT box and GATA box. The CAAT box motif (also known as CCAAT box) is a common eukaryotic motif that occurs on both strands (Roberto Mantovani, 1998). In plants the mean distance of this from the TSS is 75 bases, and it occurs on both strands (Shahmuradov et al., 2003). The consensus sequence for plants is nCAAT which was the one selected to search the chlorovirus data. CAAT elements have been found in microalgae, for example in the *C. reinhardtii* HSP70 promoter (Müller et al., 1992). GATA-boxes are the binding side of GATA transcription factors. These are present in most eukaryotes including green algae (as shown in *C. reinhardtii*) but not Stramenopiles (Thiriet-Rupert et al., 2016). In both A208R and A067R there are many CAAT and GATA motifs as seen in **Figure 5-2** and **Figure 5-3**.

Inr and DPE. The Inr (Initiator) sequence typically occurs at the transcription start site itself. The DPE (Downstream Promoter Element) is dependent on the presence of the Inr (Roy and Singer, 2015) and typically occurs +30 bases downstream of the TSS. There are several putative Inr motifs in both A208R and A067R occurring on both strands. Of these, only one will be the true transcription start site, and for A067R it seems that none of the highlighted Inr sequences may be used since they occur before the putative TATA-box and do not correspond with the approximate transcription start site from BLAST (**Figure 5-2**). For A208R one of the Inr occurs just after the TATA-box and approximately near the putative transcription start site as determined by BLAST (**Figure 5-3**). However, both these elements occur after the start codon for the gene which suggests they are either not functioning in this way or in fact that the true ORF is the second, smaller downstream one (**Figure 5-3**). There is another Inr motif overlapping with the end of the previous gene A207R which could

potentially be the transcription start site as well (**Figure 5-2**). In A067R there is no DPE but in A208R there is a DPE element on the reverse strand 42 bases downstream of the Inr which occurs after the start codon. The presence of these motifs played a part in the decision to increase the length of the chosen promoter+5'UTR sequence to include the beginning of the coding sequence of the first ORF of A208R.

Temporal motifs. Of the putative promoter regions, the presence of 'early' motifs (**Table 5-2**) further confirmed the choice of these two genes as transcriptionally early. Of the temporal motifs, A067R contained five occurrences of the early-associated ATGACAA and A208R contained three, one of which overlaps with the putative Kozak sequence around the start codon. It is interesting that A067R contains more early motifs than A208R because of the two, A208R has higher levels at the earlier time points. The promoter A067R also included a general motif of AAATTAAAA which typically occurs at -15 to -45 (Fitzgerald et al., 2008). For A067R, three out of the five early motifs are on the reverse strand and the other two are in the forward strand but occur in a tandem repeat (**Figure 5-2**).

5.2.5.4 Kozak sequence and translation start site

For translation of a protein, it is not just the start codon AUG that is important but also its context. This context is called the Kozak sequence, after a sequence bias was found in vertebrate mRNA (Kozak, 1987). The Kozak sequence is not involved in the recruitment of the ribosome (which is proposed to be via the 5'cap), but it is involved in translation initiation (Nakagawa et al., 2008). It was later discovered that there are differences in sequence preference between different species which fall into two pattern groups characterised by GCCGCCAUG and AAAAAAAUG (where AUG is the start codon), but with some common single nucleotide variations (Nakagawa et al., 2008). Consensus sequences for *C. reinhardtii* and plants are shown in **Table 5-3**.

For A067R there is a conserved plant Kozak sequence AACAAUGGC around the start codon. For A208R the Kozak-like sequence resembles the A hexanucleotide category apart from a G at the –5 position. Although C is commonly found in consensus sequences at this position (Ikeda and Miyasaka, 1998), G is not, though it

is at the -6 position. It could be that this Kozak sequence is weaker as a result. The observation of Kozak sequences does suggest that for these two genes, ribosome binding and initiation is not via an internal ribosome entry site (IRES) which are found in some viral systems.

It is worth noting that although the Kozak consensus sequences characterised in the literature include the first two or three bases after the start codon, these cannot be controlled when the part is used in a heterologous system because the coding sequence will be from a gene of interest. The choice of this codon might be restricted and might therefore not align with the requirement of the consensus. Nevertheless, it is always worth considering the Kozak consensus during the design and codon optimisation of CDS part. This is one drawback in attempting to create and characterise promoters and 5'UTRs of a certain strength, because the uncertainty of the first few bases and how this affects the efficiency of translation. This adds a source of uncharacterised variation, although through enough characterisations a predictive model could be made based on the behaviour of different coding sequences to predict the strength of translation versus the choice of second codon. Indeed, sometimes Kozak sequence priority is applied in some systems by always modifying the first N-terminal amino acid residue after the start codon to be glycine (who's codon is compatible with the Kozak consensus) as glycine is a small amino acid so is generally tolerated at the N-terminus of a protein.

5.2.5.5 The A208R upstream regulatory region was chosen to include part of the coding sequence

The putative promoter+5'UTR region of A208R has been difficult to classify due to having motifs apparently appearing downstream of the start codon as discussed above. Unlike A067R which has one ORF, A208R has three putative ORFs which are not in frame with each other (**Figure 5-3**). The first and main ORF encompasses the other two and is 936 bases (312 residues). The second ORF is located just 103 bases after the first one and is 345 bases (115 residues). The third one is still within the first ORF but towards the end and is 324 bases (108 residues).

Also A208R has several potential intron splicing donor/acceptor sites (Blanc et al., 2014).

The presence of motifs after the AUG of the first ORF, and the location of the most peroximal transcript from the BLAST analysis having a 5' end occurring after the start codon could suggest that the second ORF may be active despite it being much smaller and enclosed within the primary ORF. However, one argument against the second ORF being used is that there is no obvious Kozak around it unlike the first ORF. There is both a TATA-box and INR downstream of the first AUG, with the TATA-box ending 81 bases upstream of the second AUG, which is still further away than usual for core motifs (**Table 5-3**).

Because of this unusual arrangement, the chosen promoter+5'UTR region for A208R actually extends beyond the first start codon for 54 bases going 29 bases beyond the end of the TATA box and including the INR but not the DPE (**Figure 5-3**). At this point the sequence ends in CAAA (**Table 5-4**) and when combined with the start codon of a heterologous CDS for cloning would result in CAAAATG which potentially resembles the latter half of a plant Kozak sequence (Ikeda and Miyasaka, 1998).

5.2.5.6 Terminator region

In eukaryotes, mRNA transcription termination relies on the signal for the addition of a poly-A tail to the 3' end of the mRNA. The typical DNA motif for this is AATAAA but there are differences between evolutionary lineages, with land plants lacking specificity of the motif but tending towards a general A and T rich region, and chlorophytes such as *C. reinhardtii* having a different motif of TGTAA (Wodniok et al., 2007; Zhao et al., 2014). Downstream of the CDS for both A208R and A067R is this algal TGTAA motif (**Figure 5-2**, **Figure 5-3**, **Table 5-4**) which occurs 38 bases after the stop codon for A208R and 131 bases after stop codon in A067R (**Table 5-4**). However, it is interesting that for both genes this signal actually appears within the coding sequence of the next downstream gene since there is only 30 bases and 21 bases between the stop codon of A067R or A208R and the start codon of the next gene respectively (**Table 5-4**). Although for A208R this next gene is only a small

hypothetical reading frame of 168 bp (56 amino acids) so may not be a true gene. There are also other polyadenylation motifs, as noted in transcriptional studies (Blanc et al., 2014; Kawasaki et al., 2004) closer to the stop codons of both A208R and A067R which may be functional as well. These include two overlapping AATAAA motifs for A067R starting 7 bases after the stop codon, and ATTAAA for A208R starting 6 bases after the stop codon (Kawasaki et al., 2004) (**Table 5-4**). The approximate end of the transcript as determined by blast is just before the TGTAA signal for A208R but after the TGTAA signal for A067R (**Figure 5-2**, **Figure 5-3**). Since it is unclear which poly-A signal is being used, both putative signals were included in the chosen terminator region, so it extends into coding region of the next gene (**Table 5-4**). It is worth noting that due to the close proximity of putative genes a210L (reverse strand) A212R (forward strand) to the end of A208R (**Figure 5-3**) that there are some promoter motifs such as TATA-box occurring in the A208R terminator region. A067R terminator region does not contain TATA box (**Figure 5-2**). Neither terminator region contains the early promoter motif.

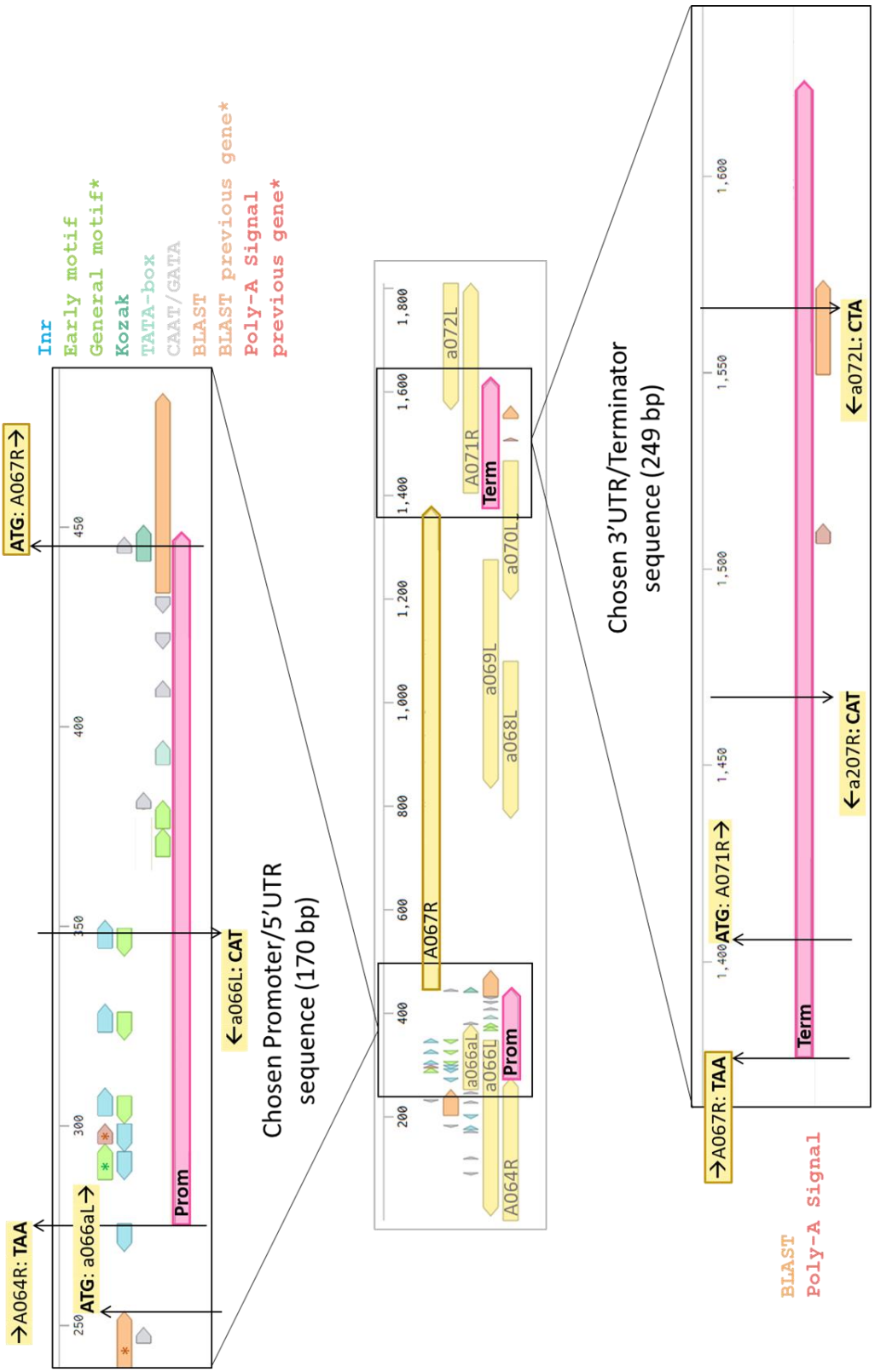

Inr
Early motif
General motif*
Kozak
TATA-box
CAAT/GATA
BLAST
BLAST previous gene*
Poly-A Signal
previous gene*
ATG: A067R
→A064R: TAA
ATG: a066aL
←a066L: CAT
Prom
Chosen Promoter/5'UTR sequence (170 bp)
A067R
A064R
a066aL
a066L
Prom
a069L
a068L
a070L
A071R
Term
a072L
Chosen 3'UTR/Terminator sequence (249 bp)
→A067R: TAA
ATG: A071R→
Term
←a207R: CAT
←a072L: CTA
BLAST
Poly-A Signal

Figure 5-2 Bioinformatics analysis of regulatory regions for chlorovirus gene A067R. Annotation of elements include the Initiator (Inr), early (AATGACA) and general (AAATTAAAA) chlorovirus-specific motifs from (Fitzgerald et al., 2008), putative Kozak sequence, TATA-box, CAAT and GATA boxes, downstream promoter element (DPE), poly-A signal (TGTAA) and approximate transcript ends as determined by BLAST analysis (see text). For a description of the precise sequences used in motif analysis please refer to **Table 5-2** and **Table 5-3**. In this case the Kozak (aacaAUGgc) mostly resembles a terrestrial plant sequence consensus (see text). In this case, due to the tight packing of the viral genome, the poly-A signal from the previous gene A064R is present within the putative promoter sequence for A067R. Multiple smaller ORFs surround, and are within, the A067R gene. A067R and A071R are the main upstream and downstream ORFs, respectively: their full length is not shown on this diagram.

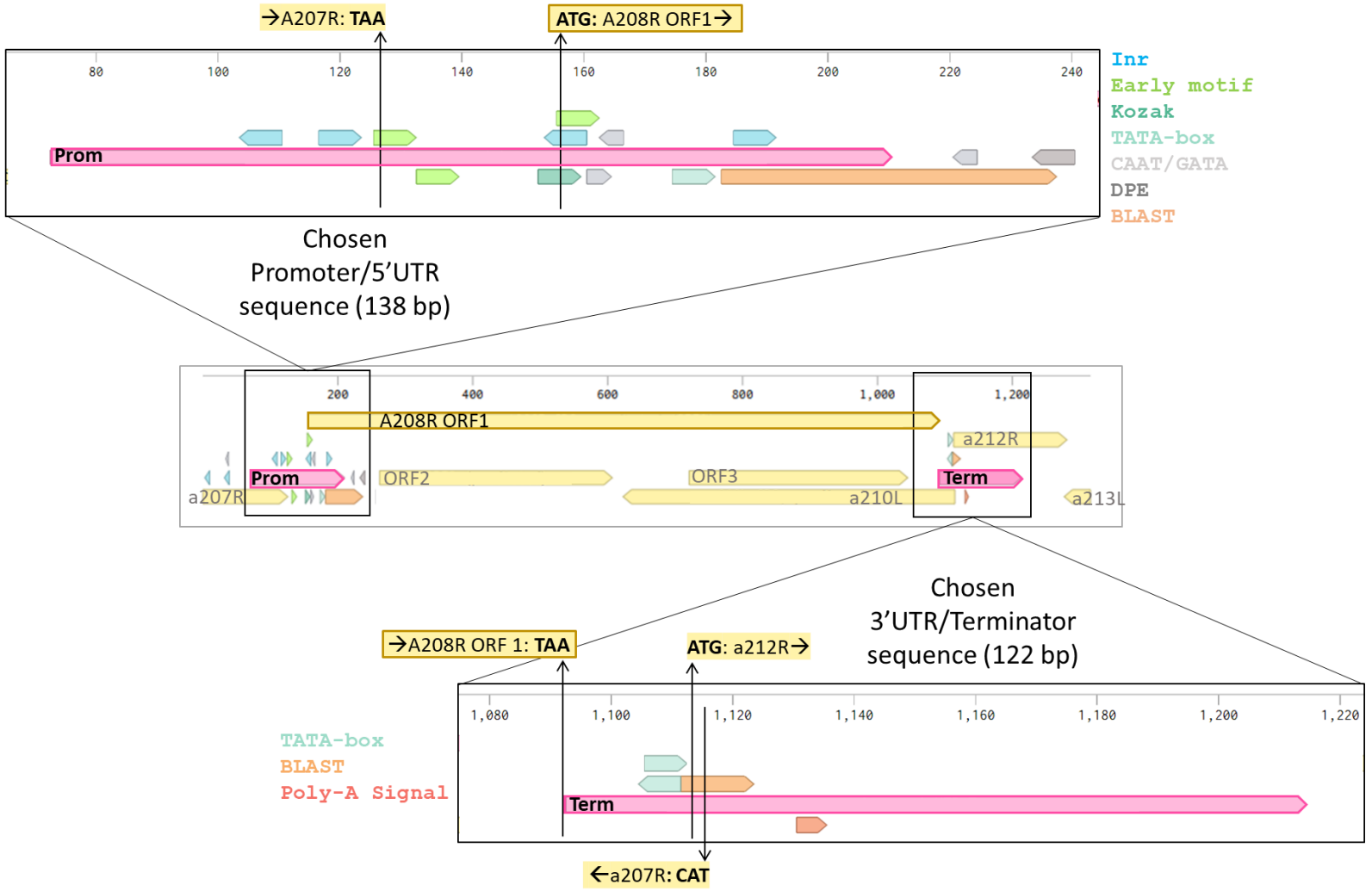

→A207R: TAA
ATG: A208R ORF1→
80
100
120
140
160
180
200
220
240
Prom
Inr
Early motif
Kozak
TATA-box
CAAT/GATA
DPE
BLAST
Chosen
Promoter/5'UTR
sequence (138 bp)
200
400
600
800
1,000
1,200
A208R ORF1
a212R
Prom
ORF2
ORF3
Term
a207R
a210L
a213L
Chosen
3'UTR/Terminator
sequence (122 bp)
→A208R ORF 1: TAA
ATG: a212R→
1,080
1,100
1,120
1,140
1,160
1,180
1,200
1,220
TATA-box
BLAST
Poly-A Signal
Term
←a207R: CAT

Figure 5-3 Bioinformatics analysis of regulatory regions for chlorovirus gene A208R. Annotation of elements include the Initiator (Inr), early (AATGACA) chlorovirus-specific motif from (Fitzgerald et al., 2008), putative Kozak sequence, TATA-box, CAAT and GATA boxes, downstream promoter element (DPE), poly-A signal (TGTAA) and approximate transcript ends as determined by BLAST analysis (see text). For a description of the precise sequences used in motif analysis please refer to **Table 5-2** and **Table 5-3**. In this case the Kozak (aaagaaaaAUG) resembles a hexanucleotide pattern (see text). Due to the tight packing of the viral genome, part of the upstream gene a207R is included within the chosen promoter sequence. Additionally, it was chosen to extend the chosen promoter sequence to beyond the putative start codn for A208R ORF1, since promoter motifs were present beyond this codon (see text).

Name	Nucleotide Sequence	Length (bp)	Description	Genomic context of chosen sequence (genes with overlapping CDS indicated by '^').
A067R Prom+ 5'UTR	5'…CGGTATGTGAAAAATTAAAA**TGTAA**TGTCATTT TACCAAGGGAGAATGTCATTTTACCAAGGGAGAATG TCATTTTACCAAGGGAGATGACAAATGACAAATGAC AATATTCAGG*TATAAA*TAAGGTGTTTTGATAGTTGA ATGTATCTTACTTATCAAAGCAACCAACA***ATG***gctc tt…3'	170	Chosen sequence is -170 bp to 0 bp of the start ***ATG***. Contains **poly-A signal**[3] of upstream gene *A064R*. Note: 3 bp (gct) after ***ATG*** form part of Kozak (dicot)[2] sequence; consider for inclusion in transgene design for optimal translation.	*A064R* (UDP-Glucose glycosyltransferase) CDS at -170 bp of ***ATG***; ^*a066L* (hypothetical protein) (reverse strand) CDS starts at -97 bp of ***ATG***; ^*a066aL* (hypothetical protein) CDS ends at -66 bp of ***ATG***.
A067R 3'UTR+ Term	5'…ta***taa***CAATTTAAATAAAATAAAATGTGAAACA TAATGACGACGAACATCCCTATTTCCCTTGGTGAAG CATACGATCGTCTTAGCATTCTAAACATCAAAAAAG AACGTTTTGTAGACCCTAAAAAACTCGAATA**TGTAA** CACGCGAAATTGAAACACTTCCGCCAGTAGAAGAAA AGTACACAGAGTTATTCAATCAGCTCCTGGATGTCA ACAAAACCCTTTGGAACGTCGAAGACAGTCTCCGTG AACTtga…3'	249	Chosen sequence begins +1 bp after A067R CDS **taa** (stop) and ends +113 bp after **poly-A signal**[3]**.**	^*a070L* (hypothetical protein) CDS begins +92 bp after **taa** (stop); ^*A071R* (hypothetical protein) CDS begins +30 bp after **taa** (stop); ^*A072L* (hypothetical protein) (reverse strand) CDS begins +191 bp after stop codon **taa** (stop).
A208R Prom+ 5'UTR	5'…aacGTACTAACAACTTCATTCAACGGTTTCGGA GAGTATGACGTGTACTACATT**TAA**ATGACAAATGAC AAATAATTAGAAAGAAAA***ATG***ACAATTGCAAAAGTG *TATAAA*TACTCTAGTCTCTATACACTTTTCATCAAA aaatcaaa…3'	138	Chosen sequence -84 bp to +54 bp of A208R start ***ATG***. Kozak-like[2] sequence resembles common A-hexanucleotide pattern.	^*A207R* (Arginine/Ornithine decarboxylase) CDS ends at -30 bp of ***ATG***; *A208R* alternative ORF (not annotated) begins at +103 bp of ***ATG***.
A208R 3'UTR+ Term	5'…ag**taa**TATATTATTAAATTATATAACATGCCAG TTAAACGACT**TGTAA**AAAGAAAAACCAATACAACGG AAAGGAAGCAAGTTGCATTGAGACAGATTCAGATAA AACATCTCAAAGCAGCGGAAAGaga…3'	122	Chosen sequence begins +1 bp after A208R CDS **taa** (stop) and ends +79 bp after **poly-A signal**[3].	^*A212R* (hypothetical protein) CDS starts +21 bp after **taa** (stop); ^*A210L* (hypothetical protein) (reverse strand) CDS begins +23 bp after **taa** (stop); *A211R* (hypothetical protein) CDS ends -47 bp before **taa** (stop) and corresponds to another ORF of A208R.

Table 5-4 Selected candidate promoter and terminator sequences from highly and early-infection transcribed genes from chlorovirus PBCV-1. The chosen sequence is in capital letters. Annotations include: early motif forward strand (solid underline), early motif reverse strand (dot underline) and general motif forward strand (wavy underline) (Fitzgerald et al., 2008); TATA Box (*Italics and dash underline*); start codon of chosen gene (***Bold Italics***), Kozak or Kozak-like sequence (grey highlight) (Joshi et al., 1997; Nakagawa et al., 2008); algal poly-A signal (**Bold**) (Zhao et al., 2014) and stop-codon (**Bold TAA**).

5.3 Discussion

5.3.1 Chlorovirus as a source of genetic parts

Inspired by the common use of genetic elements of plant viruses in plant transgenics (Govindarajulu et al., 2008), the use of algal virus parts was considered as an interesting strategy for identifying some novel DNA parts for use in *C. sorokiniana* which at present has had limited progress in transgenics. The chloroviruses were chosen because they infect species algal within the *Chlorella* genus. Specifically, the model chlorovirus PBCV-1 infects *Chlorella* NC64A. The use of DNA elements from algal viruses to drive gene expression in microalgae is not new, and the use of some of these parts previously in the literature was examined (**Table 5-2**). However, since the publication of new transcriptomic data over the course of a virus infection cycle (Blanc et al., 2014), no new promoters or DNA elements have been explored. Additionally, examination of the promoters which had previously been used showed that although the genes they selected had high transcript levels, they did not necessarily use the best gene candidates in terms of temporal expression.

Therefore, these new data were examined and two new candidates were chosen, A067R and A208R because they are accumulated in transcripts early in the chlorovirus infection cycle and are highly transcribed (Blanc et al., 2014) which suggests they are unlikely to rely on a virally encoded factor for transcription or translation. Of course, high transcription does not necessarily translate to high levels of protein expression, but it is an important part of genetic control.

One difficulty during selection of the chlorovirus DNA parts was the determination of the regulatory regions surrounding the chosen genes A208R and A067R. In the work presented in this chapter, an approach of analysing common eukaryotic promoter motifs and chlorovirus specific motifs was taken alongside looking at the transcriptomic data by BLAST. Although this technique provided a rough estimate, in the future a better way of examining this data would be to fully

assemble and align the transcriptome data, which was beyond the scope of this project at present.

5.3.2 The chosen regulatory sequences

A067R and A208R are two PBCV-1 chlorovirus genes whose regulatory sequences could be useful, based on the temporal and abundance characteristics of their transcription profile as identified from published data. Their putative promoter+5'UTR and 3'UTR+terminator regions (there was no attempt to separate the untranslated regions in this work) were analysed bioinformatically and include typical core promoter motifs, chlorovirus-specific early temporal motifs, Kozak-like sequences surrounding the start codon and putative poly-adenylation signals.

The chosen downstream regulatory sequences (3'UTR+Terminator) for both A208R and A067R included portions of the next putative CDS due to tight packing of the viral genome and the location of putative polyA signals. The upstream regulatory sequence (promoter+5'UTR) for A067R was clearer than for A208R. For A067R no sequence was included past the start codon because there were many recognisable DNA motifs in the upstream intergenic region and from BLAST the approximate 5' end of the corresponding RNA transcript was situated before the ATG (**Figure 5-2**, **Figure 5-3**). In order to include all of the relevant motifs, the chosen sequence starts the base after the stop codon of the previous gene and does include the polyA signal of the previous gene, which occurs in close proximity to a promoter motif.

For the upstream regulatory sequence of A208R the presence of motifs after the putative start codon, of a putative secondary out-of-frame ORF close to the start of the main ORF and the approximate most 5' transcript as determined by BLAST being after the start codon, led to the decision to include +54 bases of the coding sequence as part of the chosen regulatory sequence (**Figure 5-2**, **Figure 5-3**, **Table 5-4**). This may lead to problems if the first AUG, which contains a putative Kozak sequence, is recognised by the ribosome. In the classic 'ribosome scanning' model of translation initiation, skipping of any 5' AUGs is normally attributed to lack of a Kozak consensus (Kozak, 1989). However, the presence of upstream open reading frames from an non-start-codon AUG in mRNA 5'UTR is common in green algae as

characterised in the model *C. reinhardtii* (Cross, 2016), which suggests that other green algae hosts may be able to cope with this. In order to ensure there is no unwanted N-terminal extension to a cloned CDS if the first AUG was used by the ribosome, the length of the chosen sequence (54 bases) was designed to be out-of-frame when using the Golden Gate cloning strategy (which adds a four base-pair join sequence). If this sequence was to be developed and used further in the future it would be better to mutate out the upstream ATG codon and move the Kozak sequence closer to the new AUG so there is not a risk of incorrect ribosome initiation. Also, it would be interesting to know whether the inclusion of the CDS portion is needed at all by comparing different versions of the putative promoter+5'UTR which stop at the first AUG.

The use of BLAST to determine approximate transcript start and end was only a simple way of exploring the RNAseq data and hence is not necessarily accurate. If further chlorovirus genes were to be used for regulatory elements it would be better to build the full alignment of the data against the genome using a proper RNAseq pipeline as described in (Blanc et al., 2014). This strategy was beyond the scope of this project at the time, but would be useful to investigate in the future. There are increasingly more accessible pipelines available such as using the online Galaxy platform (Afgan et al., 2018) to visualise and assemble RNA data. If desired, there are ways to determine the TSS experimentally such as 5' RACE which has been used to determine the TSS of the *RBCS* gene in the microalga *Ankistrodesmus convolutus* (Thanh et al., 2012). However, since transcriptome data is available it makes more sense to do this *in silico*. The only case where the experimental approach might be more useful is if it would be desired to use regulatory regions from other homologous genes in chlorovirus who do not have this transcriptome data. This could be a useful strategy to use if there is a chlorovirus which infects a *Chlorella* sp. more closely related to *C. sorokiniana* since then the virus parts are more likely to work in the heterologous host. To properly characterise these putative regulatory sequences from A208R and A067R it would useful to compare them to the other chlorovirus ones used previously in the literature as discussed in section 5.2.2, **Table 5-2**.

5.3.3 Other potential useful or regulatory elements from chlorovirus

In conclusion, the *Chlorella* sp. viruses encode a wide array of interesting, small, unique and biotechnologically useful proteins. They also have interesting temporally controlled and highly transcribed mRNA profiles which provide a potentially valuable resource of novel promoters and other regulatory sequences for creating a toolkit for *Chlorella* sp.. Since PBVC-1 proteins are sometimes targeted to specific areas of the *Chlorella* sp. host, genes encoding these could be useful to identify putative signal sequences at either the DNA or protein level. For example, the *Chlorella* sp. virus aquaglyceroporin encoded by *aqpv1* located to the plasmalemma and plastid membranes in tobacco (Bihmidine et al., 2014). Genes containing other elements such as potential nuclear localisation sequences, internal ribosome entry sites, and efficiently spliced introns are summarised in **Table 5-5**. PBCV-1 contains novel splice sites with efficiencies up to 97% (Blanc et al., 2014; Yamada et al., 2006). Other splice sites identified are less efficient (34-45%, though frequency depended on quantitation method (Blanc et al., 2014)), and there are some apparently undesired splice sites (Blanc et al., 2014) which could indicate sequences to avoid in genetic constructs.

Gene	Notes	Reference(s)
a245r (Cu-Zn SOD)	Internal Ribosome Entry Site (IRES)	(Kang et al., 2014)
vSET	Nuclear localisation sequence; Bi-cistronic construct	(Mujtaba et al., 2008)
aqpv1	Contains plasmalemma and plastid membrane localisation sequence	(Bihmidine et al., 2014)
a125L; a185R	Contain high efficiency processed intron that is self-spliced (97% efficiency) or spliceosome-processed (89% efficiency) respectively.	(Blanc et al., 2014; Yamada et al., 2006)

Table 5-5 Other potentially useful elements or motifs from chlorovirus which may be of interest for genetic engineering of *Chlorella* sp. Cu-Zn SOD: copper-zinc superoxide dismutase; vSET: SET domain histone Lys methyltransferase; aqpv1: aquaglyceroporin.

CHAPTER 6 CREATING A LIBRARY OF PARTS CONFORMING TO STANDARD PLANT SYNTAX

6.1 Introduction

There is a need for algal biotechnology to become more aligned with the paradigms of synthetic biology including standardisation and characterisation (Scaife and Smith, 2016). By using a standard cloning strategy, parts are more easily shared between groups, and also any bespoke cloning variations and subsequent impact on DNA sequence is eliminated. Additionally, a combinatorial strategy allows for quick assembly of multiple parts to allow swapping of regulatory sequences easily and efficiently which is crucial to be able to characterise the behaviour of a DNA part in multiple settings and combinations (Patron et al., 2015). In this section, the creation of a library of parts, including the novel putative regulatory sequences from chlorovirus discussed in the previous chapter, are described.

The Chapter first describes the design and creation of small shareable library of genetic parts for *Chlorella* using standardised cloning techniques. The parts comprise promoters, terminators (including two putative promoter/terminator from chlorovirus) and coding sequences. Section 6.2 then shows preliminary evidence that the putative chlorovirus parts may be active in both *C. sorokiniana* UTEX 1230 and *C. reinhardtii*.

6.1.1 Overview of Golden Gate Cloning and the common syntax

The cloning strategy chosen is based on the plant syntax version of the Golden Gate method (Patron et al., 2015) which is ideal for assembling different combinations of parts in a high-throughput manner. The Golden Gate cloning approach uses the Type IIS restriction enzyme *BsaI* (Engler et al., 2009), which cleaves outside of its binding site to leave 4 base sticky ends. This allows for the sequence of these 4bp ends to have a defined sequence or 'syntax' for ordered assembly and ligation of fragments which is independent of the part sequence. The syntax decided by the plant community (Patron et al., 2015), which has been adopted in this work

for microalgae, is shown in **Figure 6-1**A, and divides a eukaryotic gene into twelve possible parts. To reduce complexity, if not all 12 regions are needed the spanning of parts across multiple join sites is permitted in the specification (Patron et al., 2015; Sarrion-Perdigones et al., 2011). In this work, only three parts are needed, spanning Promoter+5'UTR, coding sequence (CDS) and 3'UTR+terminator. The join sites used are those as recommended by the authors (Patron et al., 2015) and are shown in **Figure 6-1**B.

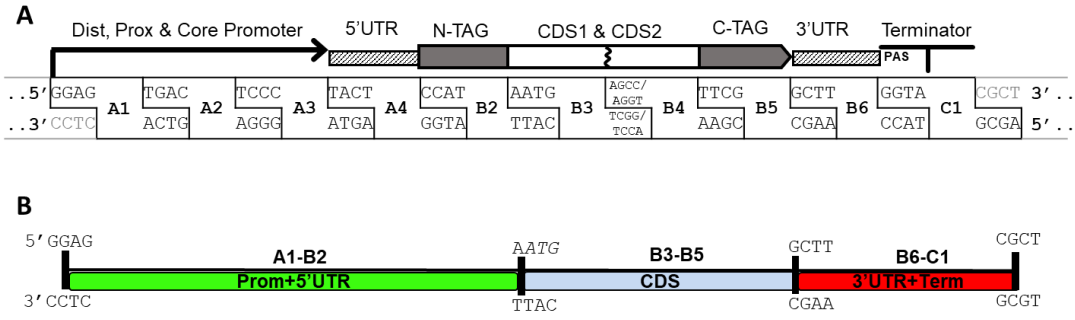


Figure 6-1 The Plant Syntax for Standard DNA Parts using Golden Gate cloning strategy. The 4 bp joins are defined by the plant syntax by Patron *et al.* (Patron et al., 2015) **A:** Description of the twelve types of standard part to include the distal (Dist), proximal (Prox) and core transcriptional promotor, the 5' un-translated region (5' UTR), a protein coding region for an N-terminal amino-acid tag (N-TAG), two coding sequences in frame separated by one of two linker amino-acids encoded within the assembly scar (CDS1, CDS2), a C-terminal protein tag region (C-TAG), a 3' un-translated region (3'UTR), and a transcriptional terminator sequence including poly-adenylation signal (PAS). **B:** Parts may span multiple join sites if not all complexity of the twelve parts are needed. In this work parts were made spanning the shown areas, with the specified 4bp junctions.

All individual parts (for example a promoter, or a reduced complexity promoter+5'UTR spanning part) are housed separately in 'Level 0' vectors which contain converging *BsaI* sites surrounding the part flanked by the desired overhangs (**Figure 6-2**A). These Level 0 parts are typically made by PCR using primers to add the relevant 4 base overhangs and *BsaI* sites. Additionally, the parts should be 'domesticated' to remove any internal *BsaI* sites to facilitate subsequent assembly. The power and ease of combinatorial cloning using Golden Gate comes from the

assembly of these parts into functioning transcriptional units in a 'Level 1' acceptor vector. This Level 1 vector must contain a different bacterial selection marker to the Level 0 parts to facilitate selection of the assembled product, and also contain diverging acceptor *BsaI* sites with overhangs corresponding to the first and last 4 base sequence in the plant syntax (**Figure 6-2**B). The Level 1 assembly is conducted in a one-pot reaction containing only small amounts of each of the Level 0 plasmids, *BsaI* enzyme, T4 DNA ligase, and the 'empty' Level 1 plasmid. The reaction is cycled between the optimum digestion temperature and the optimum ligation temperature. Of the plasmid species remaining in the reaction, the Level 0 species are eliminated due to the different antibiotic selection. The *BsaI* sites in the Level 1 vector are diverging and are therefore destroyed upon assembly completion so are not incorporated during the one-pot reaction. Any 'empty' Level 1 vector not containing the assembled transcriptional unit can be eliminated by detection using a selection screen such as blue-white selection or another visual marker such as an mRFP cassette encoding a red fluorescent protein (which is the strategy taken in this work). The completed assembled 'Level 1' vector is a functioning transcriptional unit which can be transformed into the algal host (**Figure 6-2**C).

The cloning detection system used in this work is a constitutive mRFP1 reporter cassette using the J23100 biobrick and backbone from the iGEM Registry of Standard Biological Parts (http://parts.igem.org/Part:BBa_J23100). This was added in between diverging *BsaI* sites containing sticky ends corresponding with the first and last Level 0 syntax overhangs. When the GoldenGate reaction is set up to assemble the Level 0 parts into transcriptional units, this mRFP reporter is replaced by the correct assembly and facilitates detection of correct clones as *E. coli* colonies will be white rather than red.

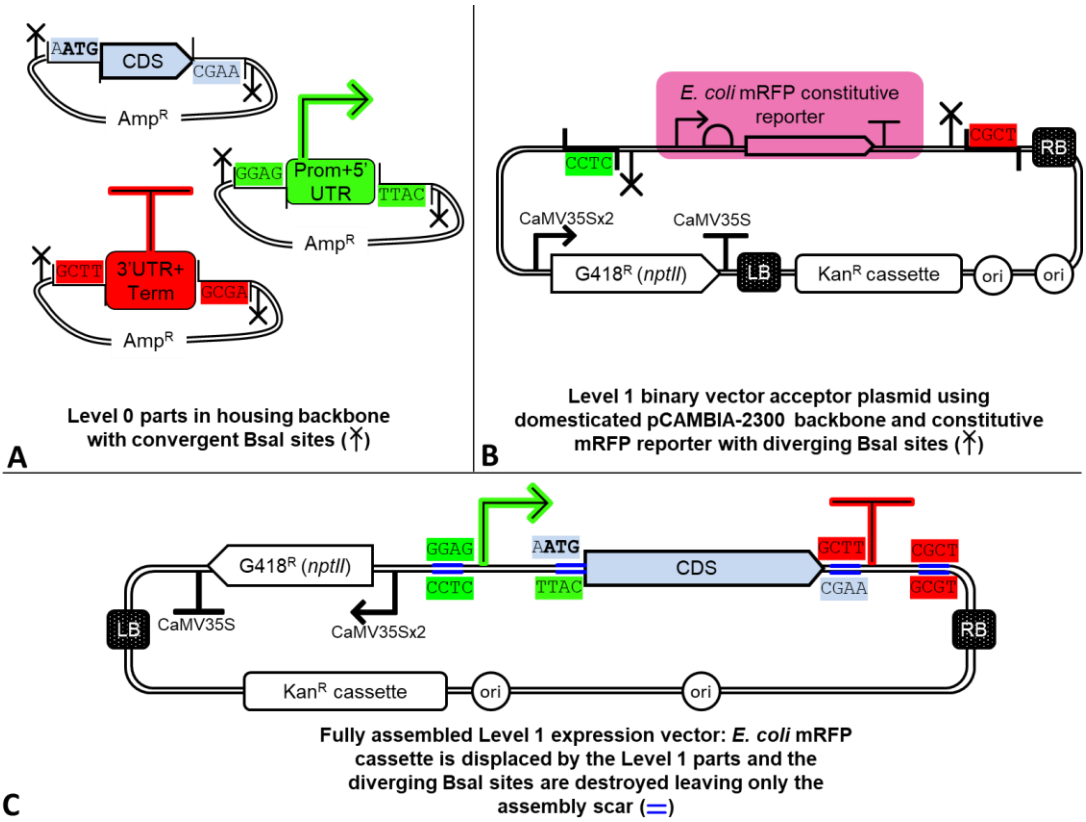


Figure 6-2 Summary of GoldenGate Cloning Strategy used in this work. A: Level 0 parts are generated and housed in individual backbones. A level 0 part is any of the individual parts described in the syntax or a span of multiple ones if not all the complexity is needed. **B:** The Level 1 'empty' vector used in this work was a pCAMBIA-2300 binary vector which has been domesticated to remove and internal *BsaI* site and modified to include an mRFP cassette as a cloning reporter (see text). **C:** The final transcriptional unit in the Level 1 vector. The binary vector backbone pCAMBIA-2300 contains a plant or algal selectable marker driven by the double enhancer Cauliflower Mosaic Virus 35S promoter conferring resistance to G418. The bacterial selectable marker is kanamycin. RB and LB are the border sequences which define the portion of DNA for transfer into the host via *Agrobacterium*-mediated transformation. There are two bacterial origins of replication, one for *E. coli* and one for *Agrobacterium*. Diagrammatic symbols used the SBOL visual standard 2.0 (Cox et al., 2018).

After assembly of parts into Level 1 transcriptional units, Golden Gate assembly can be extended further so that multiple transcriptional units can be cloned into 'Level 2' or 'multi-gene' vectors using another Type IIS restriction enzyme. The specific enzymes vary between different strategies which have been developed: *BpiI* (=*BbsI*) is used in MoClo (Weber et al., 2011), *BsmBI* in GoldenBraid (also known as GB3.0 in its most recent update) (Sarrion-Perdigones et al., 2011; Vazquez-Vilar et al., 2017) and most recently, *AarI* for Mobius assembly (Andreou and Nakayama, 2018). This layered approach makes Golden Gate a powerful tool for creating multigene assemblies which are needed for metabolic pathway insertions such as completion of the omega-3 pathway in *C. sorokiniana* UTEX1230.

The results below detail the making of the Level 0 library and assembly into Level 1 transcriptional units including the domestication of a binary vector to remove an internal *BsaI* site and use of an mRFP cassette to aid selection of correctly assembled clones.

6.1.2 Level 0 Part Library cloning

The plasmid used in this work as the Level 0 housing vector was the ampicillin resistant pJET1.2 from ThermoScientific™. This is not an official Level 0 vector but was chosen because it enables efficient ligation of blunt-ended PCR products due to its design where a cloned product disrupts a gene encoding a lethal protein termed CcdB, so it is very rare to recover empty plasmid lacking the PCR product. pJET1.2 also contains a forbidden *BsaI* site in its backbone but because the 4 base sticky end it produces is not compatible with those of the Plant Syntax, therefore there was no need to remove it. In the future it would be better to transfer these genes into an official Level 0 vector such as the 'Universal Acceptor Plasmid' pUAP2 (Patron et al., 2015).

To generate a library of parts, DNA fragments from various templates (**Table 6-1**) were amplified by PCR using specific primers (Appendix **Table A-1**) with overhangs containing *BsaI* binding sites and the relevant 4 base overhangs as previously defined (**Figure 6-1**, **Figure 6-2**A) (Patron et al., 2015). Alternately, when no template was available, the sequence was synthesised from an external company.

The PCR fragments or ordered constructs were blunt-ligated into the pJET1.2 to create the Level 0 part. The orientation of this ligation does not matter because the PCR product contains the *BsaI* and 4 base overhangs in the correct orientation with respect to the part. The complete strategy and some examples are shown in **Figure 6-3**.

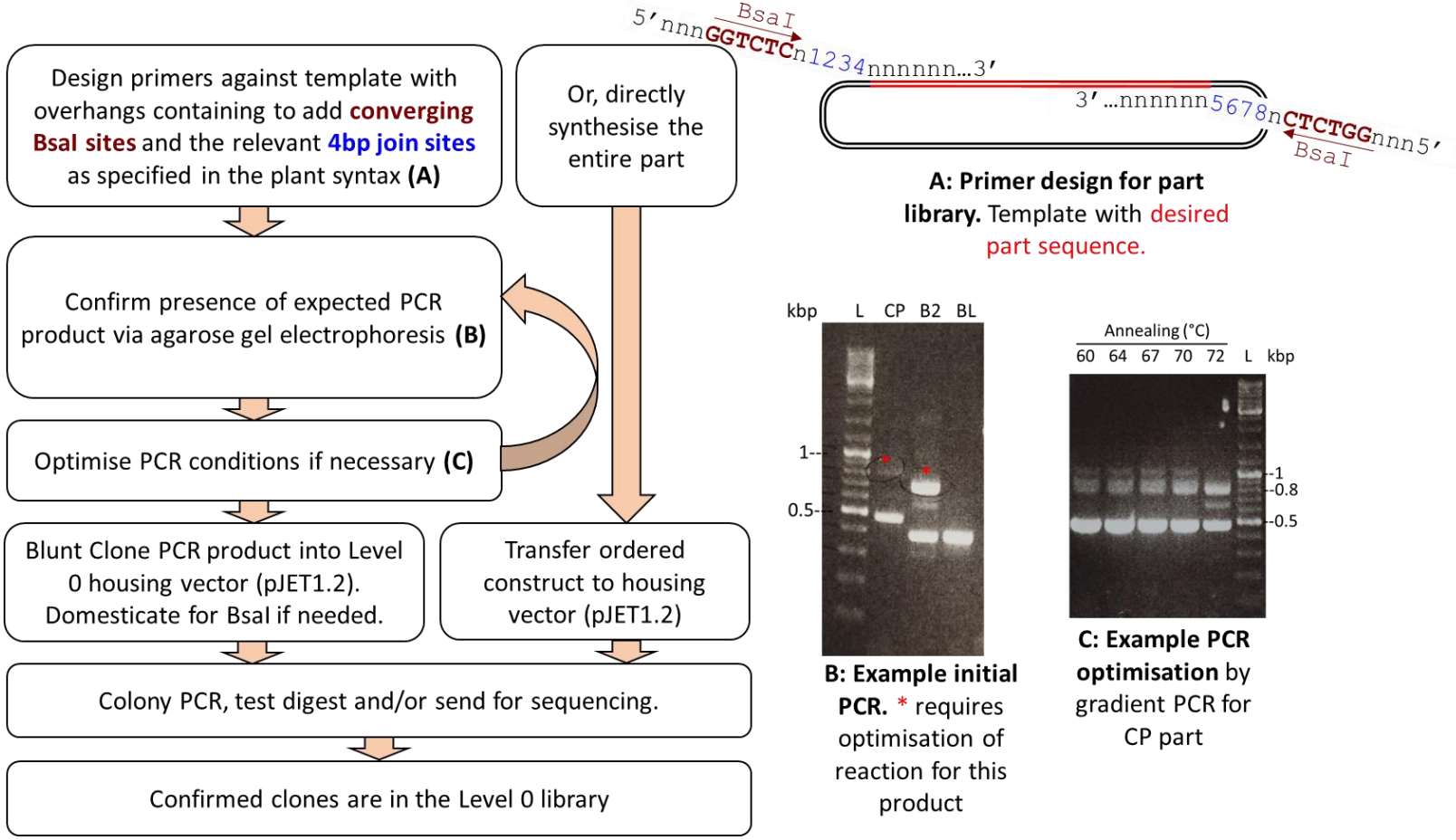


Figure 6-3 Cloning strategy flow chart and examples for Level 0 library. Gel examples include CP (CaMV25S2 promoter), B2 (Ble2I CDS) and BL (BLE CDS); see text. Ladder is Thermo GeneRuler DNA Ladder Mix.

Name	Type	Length (bp)	Primers used	DNA source or PCR template	Original source organism	Description
A2P	Prom+5'UTR	146	n/a	Ordered GeneString (A208R) or Gene (A067R) from Life Technologies	*Paramecium Bursia Chlorella virus-1* (PBCV-1)	Putative promoter+5'UTR (P) or 3'UTR+terminator (T) region from chlorovirus gene A208R (A2) or A067R (A6). The P and T for each construct are separate parts but are housed in the same Level 0 vector.
A2T	3'UTR+Term	130				
A6P	Prom+5'UTR	177				
A6T	3'UTR+Term	257				
RP	Prom+5'UTR	216	RBCSP-Bsa1-F, RBCSP-Bsa1-R	RbleC23 plasmid insert region	*Chlamydomonas reinhardtii*	Endogenous nuclear *Chlamydomonas rbcs2* (Ribulose bisphosphate carboxylase small chain) promoter+5'UTR or 3'UTR+terminator
RT	3'UTR+Term	239	RBCST-Bsa-F, RBCST-Bsa-R			
CP	Prom+5'UTR	789	CaMVP-Bsa-F(R), CaMVP-Bsa-R(F)	RbleC23 plasmid backbone region (pCAMBIA-2300)	Cauliflower Mosaic Virus	Duplicated (2x) CaMV35S promoter+5'UTR
CT	3'UTR+Term	212	CaMVT-Bsa-F(R), CaMVT-Bsa-R(F)			CaMV35S polyA 3'UTR+terminator
mCh	CDS	717	mChBsa_f, mCh-tdTo-Bsa_r	pBR9-mCherry-Cr or pBR9-tdTomato plasmid from Chlamydomonas resource centre	*Discosoma striata*	Coding sequence for the fluorescent protein mCherry (mCh) or tdTomato (td), codon optimised for *Chamydomonas reinhardtii* (Rasala et al., 2013)
td	CDS	1436	tdToBsa_f, mCh-tdTo-Bsa_r			
B2I	CDS	685	Ble2I-Bsa-F, Ble2I-Bsa-R	RbleC23 Plasmid insert region	*Streptoalloteichus hindustanus*	Coding sequence for *ble* (zeocin resistance) cassette, contains two introns from *C. reinhardtii.* (Lumbreras et al., 1998).
BL	CDS	390	BLE-Bsa-F, BLE-Bsa-R	pSP108 (Purton lab SP)	*Streptoalloteichus hindustanus*	Coding sequence for *ble* (zeocin resistance) cassette. (Stevens et al., 1996)
Od6	CDS	1387	Od6Bsa_f, Od6Bsa_r	Plasmid from Olga Sayanova (Rothamsted Research)	*Ostreococcus tauri*	Coding sequence for elongase which elongates fatty acid 18:3n3 to 20:3n3.

Table 6-1 Part source and description of the Level 0 Library. The Chlamydomonas Resource Centre is available from https://www.chlamycollection.org/. Primer sequences can be found in the appendix **Table A-1**. The GeneStrings were from LifeTechnologies. RbleC23 is an in-house plasmid made by Noreen Hiegle (Hiegle, 2014) using a pCAMBIA-2300 backbone. The length of each part includes the 4bp golden gate overlap syntax regions. The NCBI RefSeq for PBCV1 genome is NC_000852.5.

6.1.2.1 Promoters+5'UTRs and 3'UTRs+Terminators

The DNA parts representing the promoter+5'UTR and 3'UTR+terminator regions from A208R and A067R were designed *in silico* and then ordered from an external company (**Table 6-1**). For convenience, the promoters+5'UTR and 3'UTR+terminators were ordered as a single, joined construct for each gene and therefore reside in the same Level 0 plasmid. This does not cause a problem in the assemblies used in this work because there is no combining of promoters and terminators from different genes in the Level 1 assemblies. However, for future it would be beneficial to clone these parts into separate Level 0 vectors to allow for this flexibility in the assembly process.

Two other promoter sequences were included in the parts library as control promoters that have been used successfully in *Chlorella* sp. previously (see section 4.2). These are the double enhancer version of CaMV35S2 promoter and the associated terminator region as annotated on the pCAMBIA-2300 plasmid (NCBI accession number AF234315) which drives the NptII G418 selection cassette, and the *C. reinhardtii RBCS2* promoter and terminator as used in RbleC23 plasmid for testing *Agrobacterium*-mediated transformation method of *C. sorokiniana* (see section 4.2).

Of the promoters, the CaMV35S2 (CP) part proved difficult to amplify by PCR, and this is most likely due to its repetitive nature. Addressing this required optimisation of the annealing temperature using gradient PCR due to the presence of a secondary product of incorrect size. Even after subsequent PCR optimisations, the unwanted second product remained (although in reduced quantity) (**Figure 6-3**) and the correct fragment had to be purified by gel-extraction. Sequencing of the CaMV35S2 promoter revealed a few single base-pair mutations compared to the deposited NCBI sequence for the template pCAMBIA 2300 (AF234315.1). It is also worth noting that this pCAMBIA-2300 CaMV35S2 sequence as annotated on the NCBI record is markedly different to some versions of the enhanced CaMV35S promoter provided in some kits such as the Golden Braid or GB2.0 kit (Sarrion-Perdigones et al., 2013, 2011), being much longer and containing several base-pair differences (alignment can be seen in the appendix 0).

6.1.2.2 Coding Sequences

Since one of the overall aims of this project is to modify the lipid profile of *C. sorokiniana*, as a proof of concept one of the coding sequences that was included in the library is the lipid gene *Od6*. This gene encodes a delta-6 desaturase from the green microalga *Ostreococcus tauri* (NCBI accession AY746357) and is the first enzyme in the pathway to make omega-3 fatty acids from an ALA precursor (see introduction chapter 1 **Figure 1-10**). In the future, additional coding-sequence parts from fatty-acid biosynthesis genes would be added to the library. Other coding sequences that were included in the library include those for fluorescent proteins and antibiotic resistance enzymes with the aim of using them as reporters to test and characterise the regulatory elements of the library such as the new viral promoters. Fluorescent proteins are ideal because they give a graded output and lend themselves to high-throughput techniques such as using plate readers. Antibiotic resistance is useful because it too can be high throughput but can give a binary answer if detection limits for the fluorescent proteins are too low. Also, it is possible to achieve some degree of grading of phenotype by testing the resistance to different levels of antibiotic, even if not to the same qualitative extent as with fluorescent proteins. Antibiotics are also useful for cut-off selection to eliminate low expressers.

Od6. The Od6 part required domestication to remove an internal *BsaI* site. Firstly, the un-domesticated cDNA was cloned into the L0 vector in the same way as the other parts. Then, after confirmation by test digest and sequencing, the whole plasmid was used as a template for PCR with mutagenic primers (Od6-domF and Od6-domR, see **Table A-1**) to introduce a silent point mutation to remove the *BsaI* site. The mutation changes the in-frame *BsaI* binding site of GGTCTC to GGGCTC whilst keeping the coding for Gly at this position. This silent mutation of the GGT codon to GGG codon was chosen based on the codon usage table for the *O. tauri* nuclear genome as accessed through http://www.kazusa.or.jp/codon/ (Nakamura et al., 1999) which uses NCBI-GenBank 2007 data. GGT had use 24% and GGG is the next closest at 20%.

Fluorescent proteins. The fluorescent proteins chosen were mCherry (Excitation/Emission 587/610 nm) and tdTomato (Ex/Em 554/581 nm) which have been characterised in *Chlamydomonas* sp. and shown to be particularly suitable for microalgae as they were bright and their detection was not hindered by chlorophyll autofluorescence (Ex 440/9 Em 680/20 nm) which can be a problem for GFP (Rasala et al., 2013). Although in their study they report tdTomato as being the best fluorescent reporter in terms of brightness and wavelength, mCherry was also selected for this parts library because it is a monomer as opposed to the dimeric tdTomato and is and it is unknown how fluorescent proteins will perform in *Chlorella* sp. as they have not yet been used in this host. PCR gave a single band for the mCherry coding sequence but a clean amplification proved challenging for the tdTomato coding sequence and resulted in two PCR products (gel not shown), but ligation was still carried out into the Level 0 pJET housing vector and correct clones were identified by a test digest of plasmid minipreps.

Antibiotic reporters. For the antibiotic resistance reporter, the *ble* gene that encodes an antibiotic binding protein (BLE) and confers resistance to the zeocin family of antibiotics was chosen. This was because *ble* was successfully used previously in *C. sorokiniana* when testing the *Agrobacterium*-mediated transformation method (see Chapter 4). Two version of the gene were included in this library: *ble2I* and *ble*. The *ble2I* (B2) gene is the version used previously as a reporter on the RbleC23 plasmid (Section 4.3.1 **Figure 4-1**) and contains two introns from the *RBCS2* gene of *C. reinhardtii* (Lumbreras et al., 1998). When used in RbleC23 it was driven by the *C. reinhardtii RBCS2* promoter and terminator which matches the two introns. Since *C. sorokiniana* is a heterologous host and the behaviour of the introns without the *RBCS2* promoter and terminator is not characterised, another version of this gene, *ble* (BL), without the introns was also included in the library (Stevens et al., 1996). Of these two DNA parts, the B2 PCR product contained a contaminating band (**Figure 6-3**) and had to be optimised with touchdown PCR. The contaminating band was still present but ligation into the L0 vector was still carried out and then clones containing the correct product were identified via colony PCR.

6.1.2.3 Summary of completed Level 0 library

A total of four promoter+5'UTR parts, four 3'UTR+Terminator parts and five coding sequences (CDS) were cloned as DNA parts into the Level 0 vector. **Figure 6-4** lists the parts in their context as they would appear in an assembled transcriptional unit. They are all domesticated for *BsaI* so can be assembled using Golden Gate to Level 1. Assembly into multi-gene constructs may require domestication of some parts as discussed later. All parts were confirmed by sequencing except for the middle portion of tdTomato because the coding region is too large to be sequenced from the ends due to the degradation of the signal and would need to be sequenced with additional primers. However, it had been confirmed by a test digest. Additionally, during Level 1 assemblies shown later, the part was clearly functional as the tdTomato protein product was visible as an orange colour to the *E. coli* cell pellets containing the tdTomato construct.

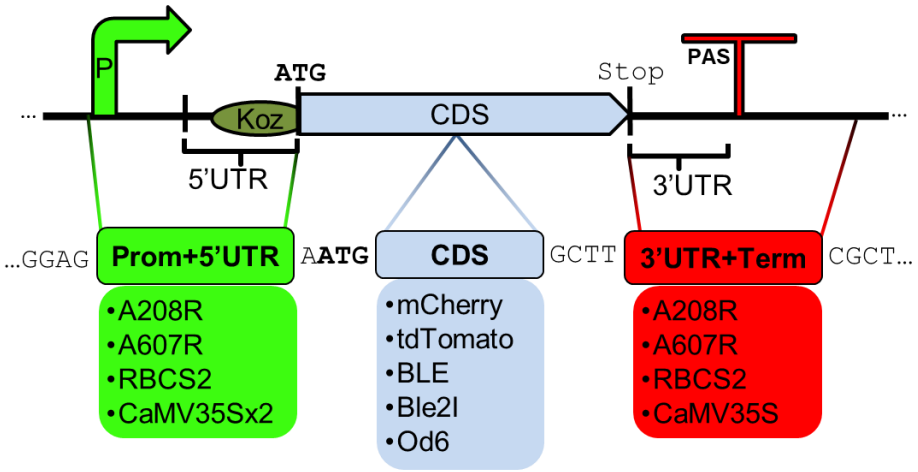


Figure 6-4 The Level 0 DNA part library. Individual Level 0 parts are shown as bullet points. There is a total of 14 level 0 parts. Their context within a transcriptional unit is shown above. P: Promoter; 5'UTR: 5' untranslated region; Koz: Kozak sequence; CDS: Coding Sequence; 3'UTR: 3' untranslated region; PAS: poly-A signal: T: Terminator.

6.1.3 Level 1 Backbone *BsaI* domestication and use of the mRFP biobrick as a cloning reporter

There are many Level 1 vectors which have been created and deposited in not-for-profit repositories such as AddGene (Engler et al., 2014; Patron et al., 2015). In this work, a Level 1 vector was made using the binary vector pCAMBIA-2300 as the backbone (http://www.cambia.org/daisy/cambia/585) because it already contains an algal selectable marker used in the previous transformation tests in Chapter 4. pCAMBIA-2300 has a bacterial kanamycin resistance gene, two origins of replication for propagation in *E. coli* and *Agrobacterium*, an *nptII* gene (G418 resistance) under control of a duplicated CaMV35S promoter/terminator as the algal selectable marker, and left/right border sequences defining region of DNA transfer to host by *Agrobacterium*. To make pCAMBIA-2300 compatible with Golden Gate cloning, it required adapting as described below. The final design, which incorporates the elements of the backbone described above, can be seen in **Figure 6-2**B. Modifications from the original pCAMBIA-2300 include:

- Domestication to remove a single *BsaI* site present in the pVS1 replication of origin
- Modification of the MCS to include diverging 'acceptor' *BsaI* sites and the appropriate 4 base plant syntax for the first and last parts in a transcriptional unit.
- The addition of a cloning reporter to aid selection colonies containing a correctly assembled product after a Golden Gate assembly reaction has been carried out.

6.1.3.1 pCAMBIA-2300 binary vector domestication (illegal *BsaI* site removal)

pCAMBIA-2300 contains an illegal *BsaI* site the pVS1 replication origin. In order to remove it the same strategy as employed by Xing et al. (Xing et al., 2014) was used to make a point mutations in the *BsaI* recognition site of GGTCTC to GGTCTG (**Figure 6-5**). A 431 bp PCR product was made using primers pCAMpVS1Bsa-fwd and pCAMpVS1Bsa-rev (**Table A-1**) which successfully amplified a 428 bp region of the

pVS1 origin of replication with the reverse primer containing the required point mutation, and also a 9 bp overhang of a *BsaI* site in the reverse orientation which would give the same cut as the original forbidden *BsaI* site. The forward primer covers an *Nhe*I restriction enzyme site already present in the backbone. The PCR product was then digested using *Nhe*I and *BsaI* and ligated back into pCAMBIA-2300 digested with the same enzymes (**Figure 6-5**). A test digest on three independent *E. coli* transformants confirmed the *BsaI* site had been successfully removed (**Figure 6-6**).

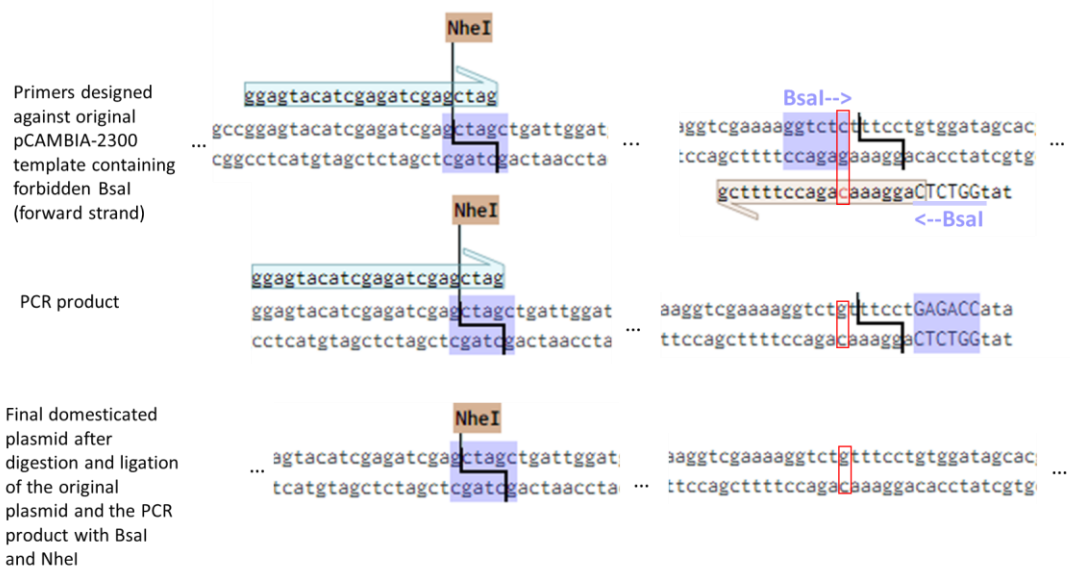


Figure 6-5 Strategy for domestication of pCAMBIA-2300 to remove internal *BsaI* site. Strategy is from Xing *et al* (Xing et al., 2014) in their domestication of pCAMBIA-based plasmids. The single base pair mutation to remove the *BsaI* site is circled in red.

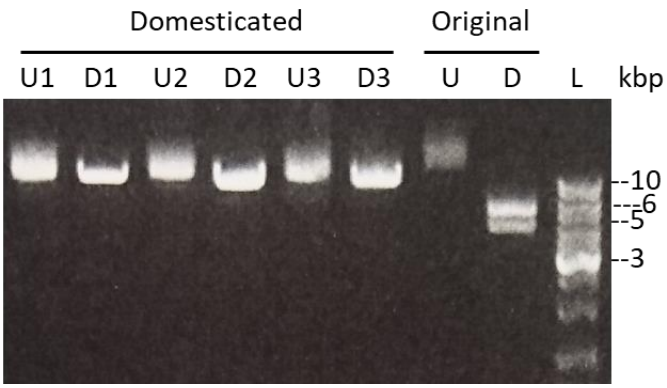


Figure 6-6 Test digest confirming successful domestication of Level 1 vector. Digests were performed with *Nhe*I and *Bsa*I which would yield a double band if both restriction sites are present (Original) and a single band if just *Nhe*I is present and *Bsa*I was successfully removed (Domesticated). Three domesticated clones were tested (1, 2, 3). Both digested (D) and undigested (U) samples were run. Ladder is Thermo GeneRuler DNA Ladder Mix.

6.1.3.2 Addition of Golden Gate acceptor *Bsa*I sites and the use of the mRFP BioBrick as a cloning reporter

After domestication the next step is to add acceptor *Bsa*I sites containing the appropriate prefix and suffix sticky ends for assembly of a transcriptional unit. This was done in the same step as adding a cloning reporter for detection of successful assemblies. The cloning reporter is required because the Golden Gate Assembly reaction may vary in efficiency depending on the size and number of parts, and there can remain some 'empty' Level 1 plasmid with no transcriptional unit present. Unlike the Level 0 plasmids, which are eliminated due to having a different bacterial selection marker, *E. coli* colonies containing this empty Level 1 vector can grow on the same selection conditions as any Level 1 vectors containing the correctly assembled transcriptional units. Although it could be possible to screen for correct assemblies via colony PCR, this would be too time consuming for high throughput parallel assemblies, especially if multiple colonies need to be screened. Therefore, an alternate visual screening method is necessary.

Blue/white selection is a common visual cloning aid and is often used, including in some of the official Golden Gate vectors deposited at AddGene (Engler et al., 2014; Werner et al., 2012), but it has disadvantages such as confining the user

to certain *E. coli* cloning strains containing the appropriate *LacZ* deletion and requiring extra reagents such as X-gal and IPTG in the plates for selection (Andreou and Nakayama, 2018; Padmanabhan et al., 2011). In this work, a constitutive *E. coli* mRFP1 cassette was used. This cassette is eliminated when a correct assembly takes its place, resulting in white positive colonies and red negative colonies containing 'empty' Level 1 vector. The use of coloured reporters has also been used in some other Golden Gate acceptor plasmids as discussed later (Andreou and Nakayama, 2018; Engler et al., 2014; Werner et al., 2012).

The source of the mRFP1 constitutive expression cassette was Biobrick J23100 from the iGEM registry (http://parts.igem.org/Part:BBa_J23100) which was kindly provided by Vitor Pinheiro lab at University College London from the 2015 iGEM distribution plate. This BioBrick is part of a collection of constitutive *E. coli* promoter variants called the 'Anderson promoter library' that is in the iGEM registry (http://parts.igem.org/Promoters/Catalog/Anderson). J23100 was shown to be the strongest promoter out of 19 variants and results in the highest expression of the product. The name J23100 technically refers to the promoter sequence only, but it comes housed in a vector containing mRFP1 and a double terminator (http://parts.igem.org/Part:BBa_J61002). A description of the mRFP1 coding sequence can be found at http://parts.igem.org/Part:BBa_E1010. The original source of mRFP1 is from Campbell *et al* (2002) but the BioBrick version has been domesticated for forbidden restriction sites used in the BioBrick RFC10 assembly standard. The cassette was successfully amplified by PCR using primers J23RFP-*BsaI*-div-F and J23RFP-div-R (**Table A-1**). These primers were designed to also add flanking diverging *BsaI* sites and the 4 base join sites corresponding the first and last parts in a transcriptional unit (A1 and C1 of the plant syntax) (**Figure 6-7**).

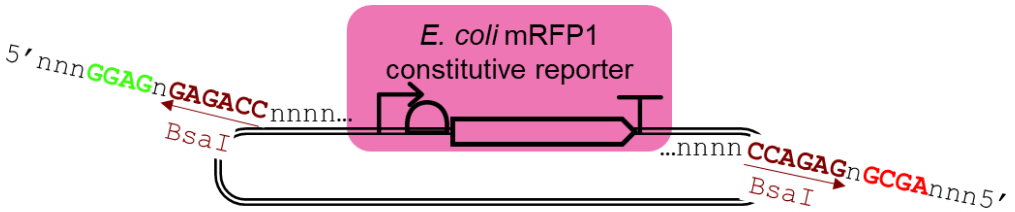


Figure 6-7 Primer design to amplify mRFP1 cloning reporter and add acceptor *Bsa*I sites for making the Level 1 acceptor plasmid. Primers contain diverging *BsaI* sites which are destroyed when Level 0 parts assemble starting and ending at the 4 base join sites. The template plasmid is BioBrick J23100 which is a promoter part housed in plasmid J61002 containing the mRFP1 expression cassette. See iGEM registry (http://parts.igem.org/) and text for more information.

This PCR product was then blunt-cloned into the MCS of the domesticated pCAMBIA-2300 plasmid via restriction enzyme *Sma*I. After transformation and plasmid DNA purification, a test digest was used to screen three colonies for orientation of the insert (**Figure 6-8**A). The colony colour was red for all three clones and the small-scale overnight culture cell pellet also shows a clear bright red colour, though with varying intensity (**Figure 6-8**B). The correct orientation was achieved in two of the three clones (**Figure 6-8**A). The correct orientation has the 4 base golden gate acceptor sites orientated so that an assembled expression construct will be in the opposite direction from the CaMV35S2-driven selectable marker already present in the pCAMBIA-2300 backbone. The completed Level 1 vector is pCor-J23-div.

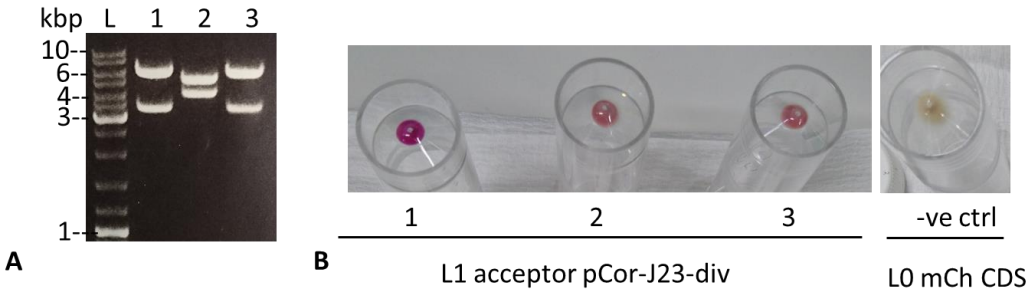


Figure 6-8 Confirmation of the mRFP cloning reporter presence and orientation. A: Test digest to screen for orientation of the mRFP1 cloning reporter and BsaI acceptor sites. The correct orientation is lanes 1 and 3. Test digest was performed with *Ssp*I and *Spe*I which cut in the plasmid backbone and at the start of the mRFP1 cassette respectively. **B:** An overnight culture of the L1 acceptor containing the cloning reporter produces a bright red product and is clearly distinguishable from cell pellet containing a plasmid without an expression cassette (L0 mCh CDS).

6.1.4 Assembly of Level 1 constructs and optimisation of the Golden Gate reaction

The assemblies were conducted in several rounds with up to ten assemblies carried out in parallel. The first round was GG1-10, the second was GG11-16, the third was GG17-20. Some assemblies required several attempts and required screening of additional colonies. During the third round, an attempt was made to alter the cycling conditions to reduce background and make the reactions more efficient. The total number of assemblies is shown in **Table 6-4** and the results and observations from examples of the assemblies and the cloning process are discussed below.

For the first round of 10 assemblies in parallel, GG1-10 were cloned. On the transformation plates there were many small colonies, the majority of which were red corresponding to 'empty' Level 1 vector. For each construct, three white colonies were picked and spotted on a copy plate to confirm absence of colour, and colony PCR was carried out for additional confirmation. The copy plate showed that some picked clones had been mistakenly identified as white because the colour was not clearly developed on the original plate (**Table 6-2**). In addition there was a potential complication with those particular constructs containing the mCherry and tdTomato parts since these genes might be expected to be expressed in *E. coli* even though they

are fused to eukaryotic regulatory elements, and would therefore give rise to red fluorescent colonies. And, indeed, this was what was seen – the PCR analysis confirmed the correct assembly despite the colonies being scored as 'red' (i.e. presumed to be the original L1 vector). This demonstrates there is background expression of the transcriptional unit in *E. coli* for some of the viral/algal/plant promoter parts. This is especially evident in GG8, which has tdTomato as CDS and the *RBCS2* promoter driving expression. GG7, which is mCherry driven by the same promoter, also shows a colour but it is less intense. Other clones in which background expression appears to occur, especially when examining an overnight culture cell pellet, is GG1 and GG2, which are mCherry driven by A2 and A6 promoters, respectively. It is interesting though that the equivalent promoters driving tdTomato (GG3 and GG4) do not give an orange product, unlike *RBCS2* (GG8) (**Table 6-2**). Finally, in some cases, white colonies unexpectedly gave a negative PCR result, for example with GG5 and GG6 (**Table 6-2**), suggesting technical failures for some PCRs.

GG rxn no.	Copy Plate colony colour			Colony PCR result correct size			Overnight culture	Overnight culture pellet colour
	a	b	c	a	b	c		
GG1	White	Faint dark red	Strong Bright red	no*	yes*	nd*	a,b	a: white b: pink
GG2	Faint dark red	White	Bright red	yes	no	nd	a	Pink
GG3	White			yes	yes	yes	a	Yellow/ white
GG4	White			yes	yes	yes	a	
GG5	White			yes	no	no	a	
GG6	Red & white	White	white	no	no	yes	c	
GG7	White	White	Pale red	yes	yes	yes	c	Pale pink
GG8	Strong orange			yes	yes	yes	a	Bright orange
GG9	White			yes*	yes*	nd*	a,b	Yellow/ white
GG10	white			yes	yes	nd	b	
-	Empty Level 1 vector pCo-J23-div: Bright red			Used as a control comparison			-	Bright red (**Figure 6-8**)

Table 6-2 Parallel assembly of 10 Golden Gate transcriptional units. This table describes round one of golden gate cloning. In the Appendix are the original images for copy plate colour (**Figure A-4**), colony PCR (**Figure A-5**) and overnight culture pellet colour (**Figure A-6**). nd: no data. A negative (No) PCR result means that the PCR product was the size of the cloning reporter rather than the size of the assembled product. *For GG1 and GG9, the assembled products have a size too similar to the size of the cloning reporter PCR product, therefore a second PCR reaction was carried out using a reverse primer (J23-mid-R, **Table A-2**) which bind only to the cloning reporter in order to determine the correct clones (data not shown).

During the second and third sessions of GG assemblies, many false positive colonies were obtained and it became evident that the high number of background red colonies was hindering the ability to pick correct white colonies. With the aim of reducing the number of background colonies containing empty Level 1 vectors, an alternate cycling protocol was tested (**Table 6-3**). This cycle was designed to include additional digestion steps and to slowly increase the temperature to the heat inactivation temperature for both enzymes because the restriction enzyme works at at higher temperatures than the ligase in the reaction, therefore reducing the amount of empty Level 1 vector because it is readily re-digested if self-ligation occurs.

Step	New (optimised) cycle		Old cycle	
	Temp	**Time**	**Temp**	**Time**
1	37	15	37	10
2	37	10	16	10
3	16	10	Repeat 3 times	
4	Repeat 2 and 3, x3 times		37	10
5	37	10	80	20
6	40	5	16	hold
7	45	5		
8	50	5		
9	55	5		
10	65	20		
11	10	Hold		

Table 6-3 Comparison of Golden Gate cycling conditions after optimisation. The new cycle was used in the third round of assembly by introducing extra steps which preferentially give better conditions for the restriction enzyme rather than the ligase in the reaction, therefore increasing the digestion of the empty Level 1 vector and reducing background on the transformation plates.

A direct comparison of the old and new cycles was carried out using the GG4 assembly which is tdTomato CDS driven by A6. The new cycle resulted in fewer background red empty level 1 vector colonies, and therefore fewer colonies overall (**Figure 6-9**). This cycle was successfully used to assembly GG17-20, which had failed to yield any correct clones in the first two attempts using the old cycle.

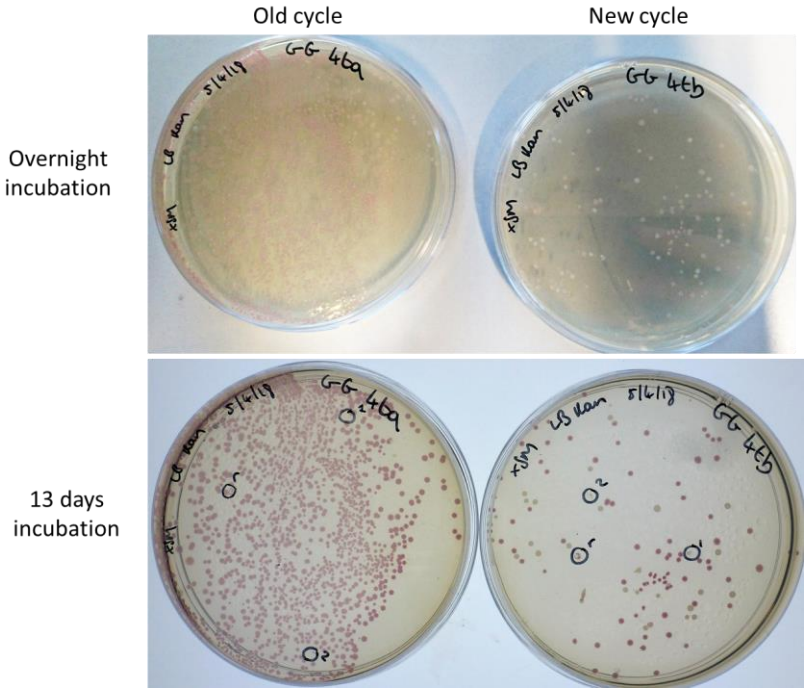


Figure 6-9 Comparison of old and new Golden Gate cycles on transformation plates. The test construct is GG4 (A6-td). The new cycle reduced the number of empty red level 1 vector colonies enabling correct (white) colonies to be seen more easily. In reality, colonies would be picked from the overnight incubation plate. For illustrative purposes only, an image of the plate after 13 days is included because the red colour accumulates more strongly over time, although in practise it would just require one day incubation and is visible on the overnight plates.

ID	Name	L0 parts used	Assembly round	Number of attempts needed
GG1	A2-mCh	A2P, mCh, A2T	1st	First attempt successful. Maximum three colonies needed to be screened for correct assembly.
GG2	A6-mCh	A6P, mCh, A2T		
GG3	A2-td	A2P, td, A2T		
GG4	A6-td	A6P, td, A6T		
GG5	CPT-mCh	CP, mCh, CT		
GG6	CPT-td	CP, td, CT		
GG7	RPT-mCh	RP, mCh, RT		
GG8	RPT-td	RP, td, RT		
GG9	A2-B2I	A2P, B2I, A2T		
GG10	A6-B2I	A6P, B2I, A6T		
GG11	CPT-B2I	CP, B2I, CT	2nd	Two attempts were needed. From 3-6 colonies had to be screened
GG12	RPT-B2I	RP, B2I, RT		
GG13	A2-BLE	A2P, BLE, A2T		
GG14	A6-BLE	A6P, BLE, A6T		
GG15	CPT-BLE	CP, BLE, CT		
GG16	RPT-BLE	RP, BLE, RT		
GG17	A2-Od6	A2P, Od6, A2T	3rd	Three attempts needed, with modified protocol. Several colonies screened.
GG18	A6-Od6	A6P, Od6, A6T		
GG19	CPT-Od6	CP, Od6, CT		
GG20	RPT-Od6	RP, Od6, RT		

Table 6-4 List of Golden Gate assembled transcriptional units in Level 1 vector. The name follows the formula of promoter/terminator-CDS. For example, for A2-mCh, the promoter and terminator are A2, and the coding sequence is mCh. Names of the Level 0 parts are explained in **Table 6-1**.

There was a general observation that for the mRFP cloning reporter to sufficiently be expressed in the colonies that plates needed to be left until the afternoon after an overnight incubation, corresponding to ~20 hours incubation. Prior to this, the colour was very faint making it hard to distinguish, but the colour continues to increase over time and is stable for several weeks. The colour was also more distinct in smaller colonies than larger ones. There was also variation in the amount of colour accumulated by some colonies. It was found that viewing the plates against a blue background aided the selection of white colonies, which were difficult to see on a grey/white background, especially if small (**Figure A-7**). However, with the optimised cloning cycle it was easy to identify the colour of colonies as there was reduced background and crowding (**Figure 6-9**).

Since the assembly reactions do not involve any *de novo* DNA synthesis, it reduces the need for sequencing. The reactions just digest DNA which has already

been sequence-verified (in the Level 0 library) and therefore confirmation by colony PCR and agarose gel electrophoresis is adequate. However, since the method was being tested, some of the assemblies were partially sequenced using the reverse primer and shown to be correct.

6.2 Beginning the validation of parts through transformation of *Chlorella sorokiniana*.

The library of assembled parts was validated and characterised by transforming the constructs into the nuclear genome of *C. sorokiniana*. Due to problems with transforming *C. sorokiniana*, the model *C. reinhardtii* was also used.

The DNA parts were validated in the assembled Level 1 constructs (**Table 6-4**). The selectable marker (in the Level 1 backbone) for all assembled constructs is G418, driven by the CaMV35S2 promoter (**Figure 6-2**). After putative transformants were obtained on primary selection plates, they were sub-cultured on selective plates. For those transformed with assemblies containing the reporter coding sequence of *shble* (GG13-16 **Table 6-4**) further spot tests were carried out on both the selected (G418) and the reporter (zeocin) antibiotic. The transformation method used for *C. sorokiniana* was *Agrobacterium*-mediated as tested in Chapter 4.

6.2.1 Transformations

6.2.1.1 Transforming *Agrobacterium* tumefaciens

A. tumefaciens was transformed with the GG constructs via a freeze-thaw protocol and made into glycerol stocks ready for use during transformation of *C. sorokiniana*. Transformed *A. tumefaciens* colonies successfully grew on plates containing the selectable antibiotic marker kanamycin. Additional confirmation of the presence of the GG binary vectors was carried out by PCR using 1-2 µl of overnight culture as template with primers pScr-fwd and pScr-rev which amplify the assembled transcriptional unit in the GG plasmids (**Table A-2**). All transformants gave a band of the expected size (**Figure 6-10**A). Some representative clones were also chosen to confirm the presence of the Ti plasmid (which contains *vir* genes essential for transfer

of T-DNA to the host) which does not contain a selectable marker and is maintained naturally by the *A. tumefaciens* strain unless it comes under stress. The presence of the Ti plasmid pTiBo542 was confirmed in all the clones tested (**Figure 6-10**B). The primers used for this screen were those used in the literature (Deeba et al., 2014) (pTiBo542-F and R, **Table A-2**). In some cases, multiplex PCR was carried out with both the pTiBo and the pScr primers in one reaction but apart from clone GG2, only the lower Ti plasmid product accumulated, reflecting the inherent issue of preferential amplification of the smaller PCR during multiplex analysis (**Figure 6-10**B).

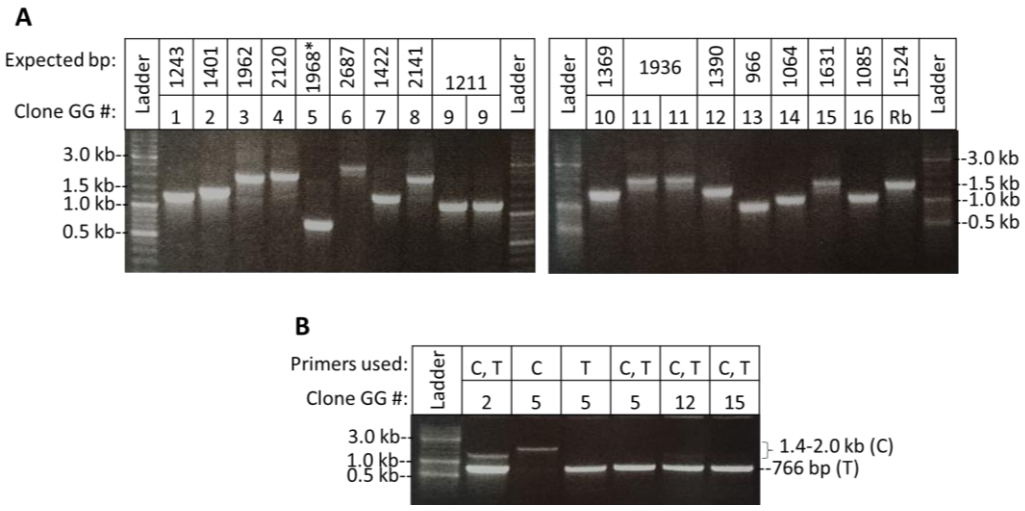


Figure 6-10 Confirmation of GG assembled binary vectors and Ti plasmid presence in transformed *A. tumefaciens*. Primer sequences can be found in **Table A-2**. **A:** Confirmation of binary vector presence by screening for presence of assembled transcriptional unit on the plasmid using primers pScr-fwd and pScr-rev. All constructs have the correct size. *in this reaction the pTiBo screening primers were also used and this product (766 bp) accumulated preferentially over the expected size from pScr. **B**: Confirmation of Ti plasmid pTiBo542 presence by PCR should give a product of 766 bp using primers pTiBo542-F and R ("T") or 1.4-2.0 kb when using primers pScr-fwd and pScr-rev ("C"). When both sets of primers are in the reaction, pTiBo product is preferentially amplified. The key to the GG clones can be found in **Table 6-4**. Rb is control plasmid RbleC23 (**Figure 4-1**).

6.2.1.2 *Agrobacterium*-mediated transformation of *Chlorella sorokiniana*

Some transformations were carried out before the establishment of the optimised transformation method (see section 4.6). The first transformation attempt

(non-optimised method) using eight different library constructs (one positive control and two negative controls) resulted in as many colonies on the two negative control plates as the putative transformants plates (Appendix **Table A-6**). Although a first propagation round was carried out on six colonies per construct on G350 selection, no further analysis was performed due to the growth of the negative control lines. These grew as well, if not better, then the putative transformants (Appendix **Figure A-8**).

Another transformation (non-optimised method 4.2) using GG5 (CPT-mCh) and GG14 (A6-BLE) and RbleC23 as a positive control only gave rise to putative colonies on the primary selection plates (G418 300 µg/ml) when algal cells was plated at a concentration higher than that determined optimal for selection. However, no colonies appeared on the negative control plate, compared to ~14-30 on the putative transformants plates. Two attempts at propagating the colonies on G418 300 µg/ml plates resulted in them dying, whereas they were able to grow on TAP containing no antibiotics. A subsequent attempt to re-introduce clones successfully growing without antibiotics to medium containing the antibiotic was unsuccessful.

Of the transformations using the optimised method (from section 4.6), there was varying success at obtaining colonies on primary selection plates. For these the assemblies chosen as test constructs were GG13 to GG16, which represent the zeocin resistance gene *ble* (with no introns) driven by A2, A6, CaMV35S and *RBCS2* promoters and terminators respectively (**Table 6-4**). These constructs were chosen because of the binary output and sensitivity of antibiotic resistance is ideal to test the ability of the putative chlorovirus parts to drive gene expression compared to the other two known promoters. The use of the *ble* gene without the introns was decided because it it not known whether the introns can be correctly and efficiently spliced in *C. sorokiniana*, or whether they interact with different promoters. Both transformations had low numbers of colonies on the primary selection plate (**Table 6-5**). Despite the method being carried out in the same way and the selectable marker being identical, different constructs resulted in different numbers of colonies. In general, survival of the first propagation on G350 was good. However, when a second

propagation was carried out on G350, more than half the putative transformants could not grow for transformation A (**Table 6-5**).

Construct	Primary selection plate colonies		1st propagation survival		2nd propagation survival	
	A	B	A	B	A	B
GG13	0	0	Nd	Nd	Nd	Nd
GG14	13	0	13	Nd	7 (4 strong)	Nd
GG15	16	0	14	Nd	3 (weak)	Nd
GG16	1	8	0	7	Nd	Nd
RbleC23	8	0	8	Nd	2 (1 strong)	Nd
-veAg	1	1	1 (weak)	Nd	0	Nd
Wt control	n/a	n/a	Unable to grow	Unable to grow	Unable to grow	Nd

Table 6-5 *Chlorella sorokiniana* colonies for two transformations on primary selection plates and their survival over two propagations on antibiotic media. The antibiotic selection was G418 at 350 µg/ml. "A" and "B" were two independent replicate transformations. -veAg is a negative control where *C. sorokiniana* was co-incubated with wild-type *Agrobacterium* during the transformation instead of *Agrobacterium* containing any of the binary vector Level 1 library constructs. Nd = no data.

6.2.2 Spot tests on putative transformants show increased resistance to the selectable marker and the reporter marker than wild-type cells

Spot tests were performed using colonies from transformation A of the new method using colonies from both the first and the second propagations (**Table 6-5**). For both propagations, the frequency of resistance to G418 was markedly lower than that to zeocin (**Table 6-6**). On the second propagation, the number of lines able to grow on zeocin declined, but for G418 the number of lines able to grow increased. All lines grew better than the escape colony or the wild-type control. The lines containing CPT-BLE (GG15) were the least able to grow on either antibiotic apart from those from the first propagation that were spotted on zeocin plates (**Table 6-6**). Lines containing *ble* driven by the chlorovirus A6 promoter were able to grow on zeocin suggesting this promoter can successfully drive gene expression in *C. sorokiniana.* The next obvious step would be to carry out molecular characterisation of these transformants via PCR and Southern blotting, however time constraints prevented this verification from being carried out.

Construct	# lines growing on Zeocin 300		# lines growing on G418 350	
	First propagation	Second propagation	First propagation	Second propagation
GG14 (A6)	10 (77%)	6 (86%)	1 (8%)	3 (43%)
GG15 (CPT)	12 (86%)	1 (33%)	1 (7%)	0 (0%)
RbleC23	3 (38%)	2 (100%)	0 (0%)	2 (100%)
-veAg	Satellites on edge	Satellites on edge	No growth	No growth
wt	Satellites on edge	Satellites on edge	No growth	No growth

Table 6-6 Number of independent putative *Chlorella sorokiniana* transformants showing growth in spot tests on zeocin and G418. The percentages are calculated from the numbers of colonies which survived propagation as seen in **Table 6-5**. -veAg is an escape colony that grew from a negative control transformation where *C. sorokiniana* was incubated with wild-type *Agrobacterium* not containing any binary vector. Wt is wild-type *C. sorokiniana.*

6.3 Validation of parts through transformation of *Chlamydomonas reinhardtii*.

After having difficulty with transforming *C. sorokiniana*, time pressure led to a decision to verify the parts in the model organism *C. reinhardtii* where nuclear transformation protocols are more well established.

Transformation was carried out using the method of vortexing cell/DNA suspension using glass beads. With the constructs chosen to be the Level 1 library plasmids containing the *ble* coding sequences driven by the four different promoters (GG13-16) since zeocin as a selection cassette is well characterised for use in *C. reinhardtii* (Lumbreras et al., 1998; Stevens et al., 1996). A key difference to the experiments using *C. sorokiniana* is that all primary selection and subsequent propagations were carried out directly on the reporter antibiotic zeocin, not G418, because zeocin sensitivity is well characterised in *C. reinhardtii*. Primary selection conditions and propagation plates were carried out with zeocin at 5-10 µg/ml.

A first transformation yielded 3-6 colonies per construct but colonies also appeared on the negative control plate (**Table 6-7**). A second transformation produced 1-4 colonies per construct including the negative control but the negative

control grew weaker over subsequent propagations. Putative colonies for GG14 and 16 also did not survive (**Table 6-7**).

Construct	**Transformation 1 picked/survived**	**Transformation 2 picked/survived**
A2-BLE (GG13)	6/6/6/6	3/2/2/2
A6-BLE (GG14)	6/4/4/4	1/0/nd
CaMV35S2 BLE (GG15)	6/0/nd	2/1/1/1
rbcs2 BLE (GG16)	3/3/3/3	4/0/nd
No DNA (-ve)	3/3(1 weak)/2/2	3/1/weak/nd

Table 6-7 Survival of putative *C. reinhardtii* colonies transformed with BLE-containing constructs from the Level 0 library. The number of colonies which survived was recorded over three propagations and is shown in the form of picked/survival propagation 1/ survival propagation 2/ survival propagation three.

Spot tests were carried out on putative transformants from transformation 2 and showed enhanced growth on zeocin compared to wild-type *C. reinhardtii* or an escape colony which grew on the negative control selective plates (**Figure 6-11**). This shows that putative transformants can be isolated using *ble* as a selectable marker for the chlorovirus promoter A208R. The survival over several rounds of propagation shows that it is not silenced in this time period.

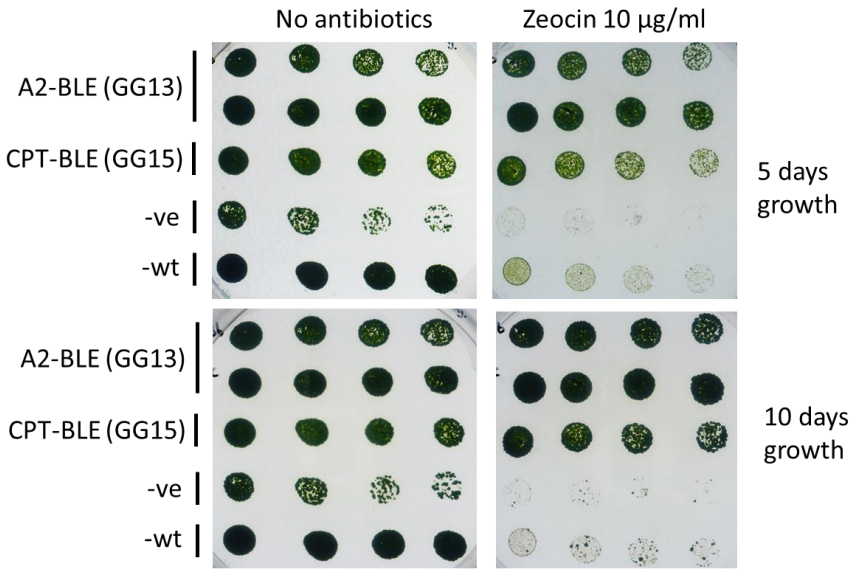


Figure 6-11 Spot tests on putative *Chlamydomonas reinhardtii* transformants containing BLE driven by chlorovirus A208R promoter and CaMV25S2 promoter. The -ve is an escape colony isolated from the primary selection plate which also survived propagation rounds. wt is wild-type *C. reinhardtii*. Cells are diluted in order of 4 spots from left to right.

6.4 A strategy to characterise relative promoter strength using the fluorescent proteins

The mCherry reporter has an excitation/emission wavelength of 587/610 nm, and for tdTomato this is 554/581 nm (Rasala, 2013). To measure relative promoter strength in the transgenic microalgae, multiple transformants will need to be screened because the *Agrobacterium*-mediated transformation method results in random integration of the translational unit into the nuclear genome. Therefore there will be positional effects impacting on gene expression depending on the genomic context surrounding the insertion site. A 96-well plate format assay would be an ideal way to screen multiple transformants in parallel. A pilot test was carried out using a fluorometer to probe the emission spectra of the proteins and assess the suitability of the available filter sets in the plate reader to confirm whether the experimental set up would be suitable. The plate reader filter sets (ex/em) were 584/620-10 for mCherry and 544/590 for tdTomato. The fluorometer was used to excite the cultures at both the optimal and the filter set wavelength, and the detection was set to scanning emission in order to observe the effects of these

excitation wavelengths. Firstly, raw *E. coli* culture expressing the proteins was measured. Then, wild-type algal culture was mixed with this *E. coli* as a preliminary way of testing whether the fluorescent signals were still visible and distinguishable in the presence of chlorophyll, which is known to interfere with fluorescent measurements of other proteins such as GFP (Rasala et al., 2013).

6.4.1 Detecting mCherry and tdTomato fluorescence using scanning emission to explore the effect of different excitation wavelengths and appropriate emission detection filters.

The constructs were expressed in an overnight *E. coli* DH5-alpha culture, cells were harvested and re-suspended in TAP medium then further diluted 1/10 with TAP. 3 ml was placed in a cuvette and measured using scanning emission at the two different excitation wavelengths in order to compare the two (**Figure 6-12**). For mCherry there was no obvious difference between the two excitation wavelengths. For tdTomato the lower excitation creates a larger difference for the emission so the peak is more clearly seen, however it is weaker than the 554 nm excitation. Also, if measured at 590 emission (plate reader) compared to the ideal 581 nm, it can be seen that for both excitation wavelengths this would give a reduced signal (left hand orange line indicates 581 nm, right hand 590 nm)

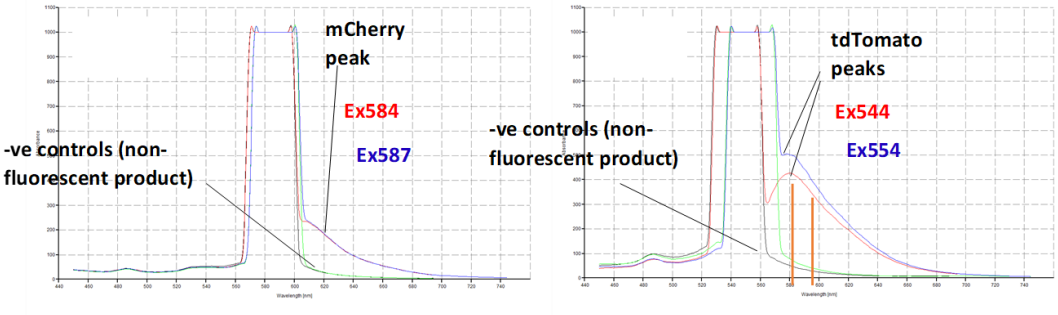


Figure 6-12 Comparing scanning emission of tdTomato and mCherry fluorescent proteins expressed in *E. coli*. The cultures were excited in a fluorimeter at two different wavelengths. The large saturated peak is the laser excitation signal. The x-axis is wavelength (nm) and begins at 440 nm, increasing in 10 nm per gridline. The y-axis is relative intensity.

6.4.2 The effect of chlorophyll from wild-type *Chlorella sorokinaina* cells on the tdTomato and mCherry signal.

To determine if the presence of the chlorophyll would interfere with the tdTomato and mCherry signals, tests were carried out where a constant concentration of fluorescent *E. coli* cells was mixed with either a high or low concentration of *C. sorokiniana* cells. The fluorimeter data (**Figure 6-13**) shows that when more algae are present in the sample, the intensity of the fluorescent signal is reduced. However, there are not corresponding changes in the chlorophyll peak.

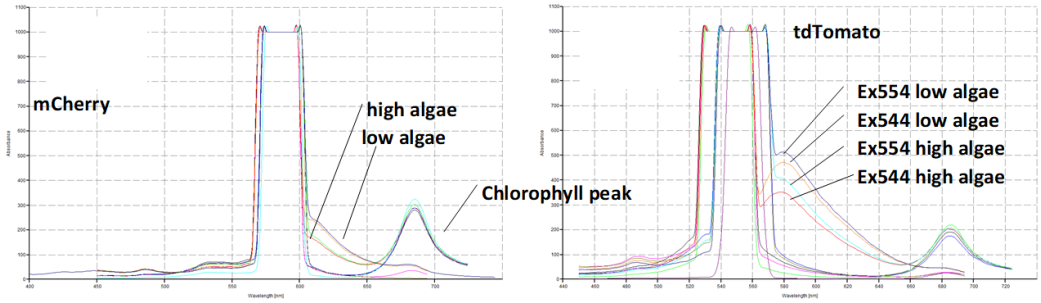


Figure 6-13 Comparing scanning emission of tdTomato and mCherry fluorescent proteins expressed in *E. coli* and mixed with two concentrations of the microalga *Chlorella sorokiniana*. The cells were excited at two different excitation wavelengths. The large saturated peak is the laser excitation signal. The x-axis is wavelength (nm) and begins at 400 nm (left) or 440 nm (right) and increases at 50 nm (left) or 20 nm (right) per gridline. The y-axis is relative intensity.

6.5 Discussion

6.5.1 A067R and A028R regulatory regions as DNA parts for *Chlorella sorokiniana* may be able to drive gene expression

In terms of assessing whether the chlorovirus promoters were able to drive gene expression in *C. sorokiniana*, it was attempted to use the selectable marker *shble* as a reporter. Low transformation efficiency, high background during selection, and the die-off of putative transformants on subsequent propagation rounds hindered the characterisation of these chlorovirus promoters and they were unable

to be properly compared to the commonly used *rbcs2* and CaMV35S2. Despite this, some putative transformants were obtained for the lines with *ble* under the control of A067R regulatory regions (GG14 A6-BLE) and were able to grow on zeocin in spot tests despite a decline in ability to grow on the selectable marker G418. This decline in phenotype in only part of the construct could be due to silencing.

An attempt was also made to test these parts in *C. reinhardtii* due to the problems with *C. sorokiniana* transformation. A preliminary transformation resulted in the generation of a putative transgenic line containing *ble* under the control of the A208R promoter (GG13 A2-BLE). Molecular characterisation of these putative transformants by PCR and Southern blot would be the obvious next step but was not carried out due to time constraints. Additional analysis could be to assess the level of transcripts by RT-PCR and a western blot to determine if translation has occurred of the protein.

For the putative *C. sorokiniana* lines, in the spot tests the cells had never been exposed to zeocin before, therefore their survival is convincing preliminary evidence that a chlorovirus promoter is working to drive gene expression in *C. sorokiniana*. The numbers of transformants obtained are too low to allow for any proper comparisons of strength between the control promoters CaMV35S2 and *RBCS2*. If the relative strengths of these promoters is to be compared to the control promoters then an average of many colonies is required because the transformation method creates lines by random integration within the nuclear genome so there will be positional effects depending on the genomic context. This would require a robust transformation method which produces high numbers of transformant colonies and a reliable selection method to reduce background escape mutants. The use of biolistics for the transformation may increase transformation rate as this is an efficient transformation method for many microalgae. For selection, alternate herbicides or antibiotics could be screened for sensitivity. Additionally, if selection continues to be an issue it could be worthwhile investing time into generating mutant strains to allow for positive selection by restoration of function such as for uracil metabolism (Kasai et al., 2015).

A preliminary test for the use of the fluorescent proteins was carried out to see if the plate reader would be suitable for detection by scanning the emission curves after excitation of fluorescent *E. coli* mixed with wild-type algae at the relevant wavelengths. Although peaks corresponding to the fluorescent emission wavelengths were visible, this test is limited because the level of expression in *E. coli* is much higher than the expression levels expected for nuclear transformation of a microalga. An alternate way to assess the relative promoter strength across a large number of independent transformants would be to harvest the total number of cells from the primary selection plate and measure the fluorescence of this mixture.

6.5.2 Cloning strategy used to make a combinatorial library

In this chapter, a library of parts was built using a well-established and standardised cloning strategy. This is relevant to algal biotechnology because as the field of algal transgenics and metabolic engineering advances there needs to be the ability for unhindered sharing of parts between labs, especially in collaborative interdisciplinary projects. Also, since transformation methods are often difficult in algae or are not well established, when creating multi-gene transgenic lines it is more desirable to perform one transformation with an entire pathway in one genomic locus rather than subsequent rounds of transformation, each one of which will be randomly integrated and need time consuming rounds of selection. The attraction of metabolic engineering to insert multigene pathways into microalgae, for example the omega-3 pathway in *C. sorokiniana* which requires an additional five genes, calls for an efficient assembly method to make these large constructs. The Golden Gate method was used here because it is well established in plants. It was successfully used in this chapter to create a Level 0 library of parts and to assemble them into Level 1 transcriptional units cloned into a Level 1 backbone that was domesticated in-house with the addition of an mRFP cloning reporter cassette. The library consists of a variety of regulatory sequences and five protein coding sequences, mainly for use in characterising the function of the promoters. Fluorescent proteins were chosen over other more established reporters such as GUS (which has been reported as successfully used in *Chlorella* sp. including *C. sorokiniana* (Liu and Chen, 2016; Yang

et al., 2016)), but the results often seem transient and not very convincing and require killing of the cells in the assaying of the GUS activity.

6.5.2.1 The mRFP cloning reporter is useful but could be improved

In terms of the Level 1 assembly process, colony PCR was carried out on multiple colonies but this would not be the aim eventually. It would be better to be confident of the cloning reporter colour by using a brighter protein or a cloning strain which is more efficient at expression foreign genes. In this work, although the coloured screen did aid in reducing the number of colonies for colony PCR, it was still quite faint and led to the picking of false positives. The reason for the faintness in colour could have been the reporter protein taking time to mature. It could also have been metabolic load because the constitute J23100 promoter chosen is the strongest in the Andersen collection. However, the colonies did not appear to be unexpectedly small or show any other signs of reduced growth.

Testing this assembly method in a different *E. coli* cloning strain might lead to higher efficiencies or easier distinction between positive and negative clones of the reporter. For example, one published study (Andreou and Nakayama, 2018) that also used coloured reporters to aid in selection of clones in a Golden Gate assembly found that the accumulation of product over time was strain dependent and that *E. coli* TOP10 accumulated faster than DH5alpha. The strain used in this chapter was DH5alpha and the time of 24 hours that they obtained in that work was similar to ~20 hours accumulation time needed in this work. The mRFP colour was adequate later in the day or several days afterwards and was very stable meaning it was ideal for coming back to cloning plates if more colonies are needed to be screened. When plates were left for at least 20 hours, the colour appeared stronger and continued to increase in strength over time. There is not a problem with temperature stability of this protein because it has been well characterised in this respect (http://parts.igem.org/Part:BBa_E1010, (Campbell et al., 2002)). One limitation to bear in mind though is that if the gene products are proteins of a red colour and there is background expression in *E. coli* then it hinders the identification of correct clones. This was the case for the Level 1 library assemblies containing mCherry. The Mobius

assembly also uses a different Anderson promoter from the BioBrick collection, J23106, which is less strong than J23100. They do not specify the reason for this choice but they do test several chromoproteins for use as their cloning reporter, including mRFP1, but end up selecting others. In the future, it could be worth using an alternate chromoprotein as the reporter. A full assessment of the efficiency of the Golden Gate reaction was not carried out. This may be valuable if larger assemblies are going to be carried out. To assess the efficiency the strategy would be to examine the ratio of assembled to empty Level 1 vectors appearing on an assembly plate under different reaction conditions.

An advantage of a standard assembly method is the transfer of parts between different lab groups and different host organisms. However, if it is going to be used across multiple hosts then species specific expression vectors for transcriptional units may be needed. Although there are many freely available Level 1 vectors in the scientific community, there may be cases where a specific vector is needed, for example if it is a binary vector or it contains some other element or selection marker. Although several Golden Gate multigene methods supply their own level vectors, the conversion of any backbone into an acceptor vector could be beneficial. This work showed the proof of concept that a cloning reporter mRFP could be used to convert a normal vector into a Level 1 acceptor vector by blunt cloning the cloning reporter after PCR with primers containing the acceptor 4 base join sites and restriction sites. The only additional consideration is domestication, but with assembly methods continually being revised and updated to include increasingly rare enzymes, this might become less needed. The addition of destination vectors for specific chassis to repositories would also allow further use and benefit the whole community.

6.5.2.2 Domestication of parts and compatibility of library for multi-gene assemblies

All Golden Gate based strategies of cloning require domestication of parts for the restriction enzyme *BsaI* and hence that is what was done for this library. For further assembly into multiple genes, the restriction enzymes used vary between different methods. Golden-Braid needs *BsmBI* and *BtgZI* which can require lots of domestication but this method and tools is continually updated and widely adopted

by the plant community (Sarrion-Perdigones et al., 2013). MoClo standard needs *BpiI* which again requires lots of domestication (Werner et al., 2012). Since both MoClo and GoldenBraid are commonly used then often efforts are made to domesticate for both *Bpi*I and *BsmB*I but although it is relatively simple with mutagenic PCR, it represents significant work and is a negative point of the methods, so much so that one of the GoldenBraid online tools helps users to find alternate sources of their gene of interest which do not require domestication. One more recent alterative is Mobius which is compatible with both eukaryotic and prokaryotic chassis and conforms to the plant syntax of overhangs (Andreou and Nakayama, 2018). Also, besides *Bsa*I the only other restriction site it needs is *Aar*I which the authors claim is rarer than the enzymes used by other standards.

In this library, none of the parts are domesticated for *BpiI* or *BsaI* which means they and their subsequent Level 1 transcriptional units cannot be used for Level 2 assemblies until this has been rectified. However, none of them contain *AarI*. The constructs containing *BpiI* are CP (6 sites), mCh (2 sites) and td (2 sites) Od6 (2 sites) and the Level 1 vector pCo-J23-div (12 sites: 6 in the CaMV25S2 promoter, 4 in the mRFP reporter cassette and 2 in the backbone). Parts containing *BsmBI* are: Od6nd (3 sites) and the Level 1 acceptor vector pCo-J23-div (2 sites). Considering that the field of standardising assemblies is progressing quickly if a new standard is adopted it may require updating. Also these assembly methods may become redundant if the cost of DNA synthesis falls lower. However, it is unlikely DNA synthesis will get so cheap that it is suitable for such combinatorial libraries. Therefore, it can be desirable to domesticate any parts for these enzymes as well. Although this was beyond the scope of this project, in the future the ability to assemble multiple transcriptional units would be a powerful tool for inserting multiple enzymes in a metabolic pathway as would be necessary for completion of the omega-3 pathway in *Chlorella sorokiniana* UTEX 1230.

6.5.3 Characterisation of the library

6.5.3.1 Problems with transformation hindered characterisation of the library

Characterisation of the library was begun using the *ble* assemblies because the antibiotic reporter was supposed to be an efficient selectable marker. Only a few clones were obtained with either *C. sorokiniana* or *C. reinhardtii*. This is not enough to properly characterise the promoters because due to the random integration into the genome of the transformation methods there are positional effects. For successful characterisation there needs to be inexcess of 50-100 clones.

The fluorescent proteins were included in the library because they lend themselves to experiments aiming to generate relative expression data in a high-throughput manner such as a 96-well plate assay. The suitability of the method to characterise the promoter strength in a 96-well plate format was tested by examining the appropriate wavelengths and testing them by mixing *E. coli* expressing the fluorescent reporters with *C. sorokiniana* to see if the chlorophyll interfered with the detection of the fluorescent signal. The result was that the mCherry and tdTomato peaks were both visible with algae therefore this is a valid strategy. Other strategies for characterising expression as an alternative to phenotypic assays if the expression level is very low which is likely so it might be below the detection limit, is to use whole plate PCR from the primary selection plates. Additionally, RT-PCR can be used to look at transcript levels and there are ways of measuring the protein.

The low transformation efficiency observed in *C. reinhardtii* was a source of concern for whether the selected lines were true transformants, since normally with nuclear glass bead transformation the efficiency is relatively high (Kindle, 1998). The plasmid was relatively large (over 9 kb) because the Level 1 vector is designed for *Agrobacterium*-mediated transformation, which could have reduced efficiency. However, the spot test of putative transformants driven by A208R chlorovirus promoter on zeocin did suggest they are true transformants. But to be confident, molecular confirmation such as genomic PCR or Southern blot would be required. A good next step would be to test these lines on G418 since it is also on the construct as the selectable marker.

For *C. sorokiniana,* problems with the *Agrobacterium*-mediated transformation could have occurred for many reasons. As discussed previously in Chapter 4, the method was hard to optimise. Additionally, there could have been a problem with the *A. tumefaciens* strains. For example, there is variation between transformation repeats as to which constructs produce transgenic lines, despite the selectable markers being identical. However, caution must be taken making conclusions for this because the numbers are very low and the differences may be due to chance. Gene silencing is also a common problem in microalgal transgenics but the mechanisms are not well characterised. The loss of resistance to the selectable marker G418 in some lines suggests that some form of silencing may be occurring in *C. sorokiniana*.

CHAPTER 7 FINAL DISCUSSION

7.1 Introduction

Over the past two decades many microalgal species have received positive attention in the scientific community, and typically any one report on an organism will state that it is of unique industrial relevance. *C. sorokiniana* is one of these such strains for reasons as described previously, namely its fast growth at high temperatures and reports of lipid accumulation under growth conditions containing exogenous carbon sources.

The work presented in this thesis revolves around the study of the behaviour of a complex biological system (an algal cell) in response to several perturbations: the aspects investigated, such as lipid profile, growth, and genetic transformation, depend upon central metabolism which is inherently an intricate multivariate network with a multitude of feedback loops. Attention was drawn to this in the introduction where it was highlighted that although several factors are known to influence lipid profile of microalgae, the precise behaviour of the cell in response to perturbation of these factors can be unpredictable, even within the same strain, because minor differences in environment or cell state can cause interference. Also, crucially, the observations are not consistent across the highly evolutionary and physiologically diverse polyphyletic microalgae. These unpredictable effects are not limited to studies on lipid accumulation in Chapter 3, but also impacted the work in Chapters 4-6, as seen in the high variability and inconsistency in antibiotic sensitivity, where cells are exposed to physiological and environmental stress during transformation and have uncharacterised cellular mechanisms for coping with heterologous DNA within the cell.

Therefore, there are some conflicting results which may raise more questions than answers. It is the intention here to provide some insight into any confusing data which arose during these investigations and suggest ways of addressing this in future work. In this chapter, the thesis findings are summarised, before discussion of the

short term experimental future work and particularly interesting results. Finally, the consideration of this work within its wider context is presented in an outward-looking manner alongside some potentially fertile avenues for future research in algal biotechnology.

7.2 Summary of main aims and findings

The aim of this thesis was to explore the potential of *C. sorokiniana* UTEX 1230 as a production platform for high value oils, including the proof-of-concept genetic engineering which would be required to extend the wild-type omega-3 pathway from terminating at ALA (C18:3n3) to synthesis of the LC-PUFAs EPA (20:5n3) and DHA (22:6n3). Therefore, work in this thesis includes investigating *C. sorokiniana* lipid accumulation (chapter 3), a method for transformation and selection (chapter 4), development of genetic parts required to drive gene expression (chapter 5 and 6), and construction of a platform and parts library for rapid gene assembly (chapter 6).

In Chapter 3, the bulk of work was concerned with assessing growth and lipid profiles in different nutrient regimes in order to determine suitable conditions for the accumulation of the omega-3 precursor C18:3n3. Differences in the FAME profile were observed in response to duration of cultivation, trophic mode and base media. A screen of fourteen carbon and four nitrogen regimes for cells grown in mixotrophy after 48 hours found that when grown on 17.4 mM acetate cells had the highest proportion of C18:3n3 (36 mol % total lipids), but 500 mM glucose had the highest C18:3n3 by mass (0.38 pg/cell), and 17.4 mM acetate + 5.8 mM glucose resulted in the highest C18:3n3 productivity (9.8 µg/ml culture). A preliminary assay confirmed that nitrogen stress and high glucose concentration are triggers for neutral lipid accumulation, but found that the resulting shift in the FAME profile meant that on a whole culture productivity level, in general replete nitrogen had more total C18:3n3 across all lipid after 48 hours growth. The shift of the FAME profile to saturated fatty acids under nitrogen stress suggests that the use of this well-known strategy is suited to biofuel applications but could be counter-productive for PUFA accumulation. It was also found that cultures with a low pH show an observable increase in neutral

lipid without a reduced proportion of C18:3n3 in the fatty-acid profile, though more rigorous analysis is needed to confirm this data.

In Chapter 4, efforts were made to build upon previous work using *Agrobacterium*-mediated transformation for the nuclear genome of *C. sorokiniana*. After tests using a prior protocol for *Agrobacterium*-mediated transformation highlighted inconsistencies in the efficacy of selection, extensive antibiotic sensitivity assays were carried out for G418 and Zeocin™. These studies found that selection using these antibiotics was sensitive to small changes in cell plating density and that false positives were common. However, ultimately a concentration 350 µg G418 at 100 million cells per plate in soft agar was found to be most effective. Following which, efforts were focussed on improving the efficiency of the *Agrobacterium*-mediated method. After surveying the literature on other variations of this method used in other microalgal species, and combining insight from the infection mechanism of *Agrobacterium*, multiple points for potential improvement of the method were identified and tested. These included additions of extra steps for *Agrobacterium* virulence induction, more careful control of algal cell numbers to align with the selection strategy, and the removal of an overnight recovery step. The data suggests these are promising modifications but more rigorous direct comparisons wit the orginal method are needed to confirm their precise effect.

In Chapter 5, an investigation into the potential of the use of chloroviruses as a source of genetic parts for driving gene expression in *C. sorokiniana* was carried out. Two genes, A208R and A067R, which are highly transcribed in the early infection cycle of chlorovirus, were chosen and their putative regulatory regions were bioinformatically analysed to identify the presence of several promoter and terminator DNA motifs and elements which informed the choice of the precise portion of the sequence to use as the DNA part. Preliminary evidence for the ability of these parts to drive gene expression in *C. sorokiniana* and *C. reinhardtii* was shown in Chapter 6: putative Zeocin™ resistant lines were obtained for *shble* (without introns) driven by A067R for *C. sorokiniana* and A208R for *C. reinhardtii*, though confirmation at the molecular level is still needed.

In Chapter 6, a library of DNA parts compatible with the Golden Gate cloning strategy and conforming to the established plant syntax for defined joins between parts was built. The Level 0 library is domesticated for the restriction enzyme *BsaI* and consists of four promoters+5'UTRs, four 3'UTR+terminators and five coding sequences. The promoters and terminators include the commonly used CaMV35S2 and *RBCS2* as standard parts, and those sequences newly isolated from the chlorovirus genes A208R and A067R, which to the authors knowledge have not been previously used. The coding sequences are two fluorescent proteins (tdTomato and mCherry), two versions of the *shble* gene (with and without introns) which confers Zeocin® resistance, and the *Od6* desaturase which is the first enzyme in the omega-3 pathway. These Level 0 parts were assembled into 20 Level 1 transcriptional units using a domesticated binary vector backbone which had been modified to include a chromoprotein as a negative cloning reporter. The assembly process and cloning reporter was well validated in this chapter, though some improvements are suggested. The library is available for sharing with other groups but requires further restriction enzyme domestication depending on the precise Type-II restriction enzymes used in further assemblies, of which there are multiple variants.

7.3 Discussion and short term future work

7.3.1 Evaluating nutrient regimes for lipid accumulation

Even though some previous investigation on the impact of trophic mode and carbon source of *C. sorokiniana* lipids have been reported as described in Chapter 1, the high variation in microalgal lipid work means that ultimately it is valuable to know the effects are reproducible in a different laboratory environment. Additionally, there were specific questions related to the aim of this thesis that this chapter aimed to address which were not already considered in these previous reports. Specifically, this work was concerned with the effect of these perturbations on the abundance of the fatty acid C18:3n3.

When using the different base media BBM and TAP, although it is logical to assume that the variances in growth and FAME could be due to the different carbon

and nitrogen sources in these two media, examination of the molar elemental composition of these two media highlights differences in ions such as phosphate and sulphate and trace metal solutions and iron which are known to be potential influencers of lipid metabolism. Therefore, these may also play a part and their influence could be a source of further investigation. An optimal way to perform this would be to use a systematic Design of Experiments approach using high-throughput technologies which have been previously used to successfully optimise nutrient requirements in microalgae (Hallenbeck et al., 2015a; Radzun et al., 2015; Wolf et al., 2015). In fact, authors of these multivariate analyses highlight that consideration of a single nutrient factor at a time misses hidden effects, and that unconsidered changes in ionic content which occur whilst changing specific elements of the media, can confound results (Evens and Niedz, 2010).

High throughput and robotic approaches have the potential to greatly enhance the speed and rigour of many microalgal studies, and attribute proper significance values to observed effects. In this work the presentation of the results did not contain any statistical tests because it was beyond the scope of these preliminary screens. Therefore, the observations were just correlational, and analysis of the data was relatively qualitative. However, useful avenues for potential future avenues which could confirm these associations were identified as described below.

It is essential to confirm the association of acidic pH and neutral lipid accumulation with additional experiments because the observations presented could have been confounded by multiple factors. These hidden interactions could include cell size, the unreliability of the Nile red assay in terms of convoluted raw curves, and the impact of the media pH on the Nile red binding reaction itself. The experiments which could be carried to verify the acidic pH could be conducted by rapidly reducing the pH of healthy cultures and measuring the effect on total lipid, neutral lipid and FAME. Other useful experiments involve the use of labelled carbon to monitor the rate of de novo fatty acid synthesis in these different conditions and give insight into cell remodelling and fatty acid transfer between pools. If acidity were to be an effective lipid accumulation trigger, it could be relevant to industrial lipid production in microalgae because it would involve the cheaper process of addition of a

compound rather than removal or change of media. Additionally, it would allow the use of nitrogen rich nutrient sources such as those of waste sources, which are rich in organic nitrates and phosphates and so would improve the economics by not relying on defined media. Additionally, an alternative to nitrogen stress cultivation would be more suited to enrichment of high-value LC-PUFAs because it avoids the shift in FAME profile to more unsaturated fatty acids.

These data suggest that out of the conditions screened, although nitrogen deprivation has higher lipid productivity than replete for some fatty acid species, the use of exogenous carbon sources is a better strategy for PUFA (such as 18:3n3) productivity. For N-deprived cultures, the fatty-acid profile seems to suit biofuel applications more, with a preferential accumulation of saturated, mono, or di unsaturated fatty acids. The suitability of the high concentration (9% w/v) of glucose in this experiment was unable to match the productivity of the lower acetate and acetate+glucose cultures, but it did have a more promising mol% profile for 18:3n3 and examination of FAME profile of this condition over a time-course would be a useful further investigation.

This shift in FAME distribution between regimes is another interesting point for further research and it would be interesting to explore whether the shifts are in certain lipid classes. This could be tested by quantifying the difference of FAME in TAG against total FAME, or characterising FAME in each lipid class separately by techniques such as mass spectrometry. The desire to have certain lipid species in TAG compared to other lipid classes depends on the application. For nutritional purposes, if the final product is to be total biomass as feed then in fact the presence of omega-3s in forms such as phospholipids may be better than TAGs because they are more bioavailable in this form.

For exploration of carbon and nitrogen regimes, the next valuable experiment would be to take the top performing conditions for C18:3n3 and combine them. For example, taking the top two performing carbon sources, which are Ac and AcGlc, and testing these with the two different nitrogen sources. Testing higher concentrations of the sugars to a level where they do not kill the cells would be interesting to see if

they supported higher cell densities. Additionally, performing this in heterotrophy instead of mixotrophy would be valuable to know whether the presence of light was beneficial or a hindrance to the growth of the microalga. Essentially, the screening of the carbon and nitrogen source analysis performed provides us with a snapshot at 48 hours cultivation which is unlikely to be the optimal harvest point for each growth condition. An experiment performing these lipid analyses over a time course whilst monitoring the growth curve would confirm the growth phase of the cells and allow true productivity in terms of lipid amount per culture volume per day, which is necessary for consideration in industrial settings.

True in-depth analysis of such a complex adaptive system would require a proper Design of Experiment approach with the appropriate statistical tests such as ANOVA and ANCOVA. These type of experiments have been carried out in some microalgal research but the number of such publications is small (Chiranjeevi and Venkata Mohan, 2016; Hallenbeck et al., 2015a, 2015b; Wolf et al., 2015). This type of experiment will be a powerful tool for probing the vastly complex multivariate network of central carbon metabolism. Metabolic models of multiple microalgae now exist including a *Chlorella* species (Huelsman, 2015; Zuñiga et al., 2016), but their use in terms of informing experimental design appears limited and at present they remain an academic exercise despite having value if people were confident enough to use and explore their use in industry. True integration of *in silico* modelling to inform experimental design and hypotheses is a crucial way to examine multifaceted problems, but requires an integrated multidisciplinary team.

Ultimately, the transformation strategies and genetic parts such as those developed in chapter 4-6 could be used to probe the system by over or under expressing certain genes and examining andy resulting changes in lipid profile across acyl chain length and lipid class.

7.3.2 Transformation of *C. sorokiniana*

This project was critical to the overall aim of creating a transgenic line of *C. sorokiniana* containing omega-3. Although some putative clones were obtained using both the original and the modified method, it is unclear as to which has higher

efficiency because the variability was so high and a direct comparison was not carried out. Additionally, there were problems in a decrease in phenotype over time. Despite the difficulty in getting transformation working in *C. sorokiniana*, *Agrobacterium*-mediated transformation is still a valid strategy to pursue for microalgae biotechnology, especially the application of transforming multiple genes in one locus, because *Agrobacterium*-mediated transformation method is reported to transfer up to 100 kb successfully.

It would be beneficial to repeat phenotypic and molecular assays on old transformant lines as it will be valuable to know how stable the transformant lines are in the long-term. A direct comparison of old and new *Agrobacterium*-transformation method would inform as to whether this rational method of modification is a useful strategy and further areas for improvement include optimising the ratio of *Agrobacterium* to *C. sorokiniana* cell ratio and maximising the surface contact area of the *Agrobacterium* to *C. sorokiniana* cells using novel agar plate shapes. It would also be good to properly monitor the stability by maintaining the same transgenic line on both selective and non-selective media and periodically monitoring both phenotype and molecular confirmations.

7.3.3 Genetic parts from chlorovirus

Inspired by the success of regulatory sequences derived from plant viruses to drive high levels of gene expression in plants, an investigation into the potential of the use of chloroviruses to be a source of parts for driving gene expression in *C. sorokiniana* (or other green algae) was carried out. Clearly molecular confirmation is the next essential step, with PCR being the most direct method of achieving this. Southern blot could then be used to determine integration pattern and copy number. Transcription and translation would then be examined by RT-PCR and western blot respectively.

There were some puzzling observations during the annotations of promoter motifs in the chlorovirus genes such as the uncertainty of the correct start codon for A208R gene. Viral genomes often have unusual genome structures so it could be that there is some unidentified mechanisms and cryptic binding sites for transcriptional

or translational control factors. Also, the genome is very densely packed which can lead to complex multifunctional motifs being present and cryptic binding sites for DNA-binding factors. One way to test the functionality of the putative promoter elements would be to test versions of the promoter that are sequentially shorter.

Proper mining of chlorovirus genome using algorithms and the proper assembled transcriptome might identify other novel gene expression elements that could be added to the parts library. Alternatively, one could take the high throughput approach of creating a library from genomic preparations of chlorovirus DNA and testing whether they can drive gene expression. However, this would rely on having access to a high-throughput and reliable transformation method.

Comparison of promoter strength of the putative chlorovirus promoters against the standard promoters could be carried out either in 96-well plate format for single colony isolation or pooling of a whole set of clones from a transformation plate. A large number of clones needs to be screened when assaying promoter strength so to average out the positional effects seen as a result of random nuclear integration. Pursuing strategies such as the use of DNA editing technologies to ensure insertion into the same position in the nuclear genome then screening this number of independent transformants would not be necessary.

7.3.4 Library of parts

An increasing shift towards a synthetic biology approach revolving around a design-build-test cycle requires standard parts which are well characterised. To be well characterised, a part must be validated multiple times in different conditions, and in order to do this these parts need to be shareable between labs. This also enhances the progress of algal biotechnology as it is easier to adopt and use parts if they are useable straight away rather than having to convert them into a bespoke cloning strategy.

This library is relevant to algal biotechnology because there is increased interest in metabolic engineering approaches and tuning expression via promoters of various strength would be useful. Additionally, reconstitution of metabolic pathways

in heterologous organism may require several genes and there needs to be a suitable cloning strategy for coping with this. For *C. sorokiniana* to make tailored oils such as EPA or DHA it requires the insertion of three to five heterologous genes – a considerable operation by conventional cloning.

The chosen fluorescent protein reporters have the potential to be more useful than others such as GFP because they were deliberately chosen as ones that were characterised in algae to have low interference from chlorophyll a and b (Rasala et al., 2013). The parts library cloning system is functional, and the DNA can be freely given to other groups. However, it would be useful to conduct the domestication of the parts in the library for subsequent Level 2 assembly restriction enzymes depending on the desired Level-2 standard to use. The addition of further fatty acid elongase and desaturates from the omega-3 pathway to the Level 0 parts would enhance the variety of genes in the library. A good validation of the cloning reporter system would be to conduct a multigene assembly into Level 2 units to include the transcriptional units of the 5-genes required for the pathway for omega-3 LC-PUFA DHA.

7.4 Ideas for longer term progress of the field of algal biotechnology

7.4.1 The potential to take advantage of the promiscuity of *Agrobacterium* to act as a multi-species transfer agent of proteins and DNA

The use of *A. tumefaciens* to transfer and integrate T-DNA into the host genome is an important technology. However, there is also the potential to exploit the *A. tumefaciens* method to transfer proteins to the algae in addition, or instead of, the T-DNA. This has exciting possibilities to facilitate the targeted modification of nuclear genome using genome editing technologies, a highly desirable and powerful tool in which algal biotechnology is lacking.

Transfer of proteins is possible by fusing a type-VI secretion signal tag to the protein (Vergunst et al., 2005) and has been employed in yeast to transfer a nuclease to create double stranded DNA breaks and increase the probability of DNA repair by homologous recombination (Rolloos et al., 2015). Perhaps in extension to this,

nucleases used in genome editing systems such as the Cas9 of the CRISPR/Cas system could be transferred. CRISPR/Cas9 has been successfully employed in the diatom *P. tricornitum* (Nymark et al., 2016), but not in *C. reinhardtii* where the Cas9 is toxic when constitutively expressed (although it is shown to be functional). The authors of the report emphasised the need to stop Cas9 production or activity after gene editing (Jiang et al., 2014). The only way CRISPR/Cas9 (or the similar Cpf1) has been shown to work in *C. reinhardtii* is by transfer of Cas9 or Cpf1 ribonucleoproteins (RNPs) rather than expression of the nucleases (Ferenczi et al., 2017; Shin et al., 2016).

It is likely that other green algae may also be intolerant of Cas9 expression, and additionally, many microalgae remain recalcitrant to transformation techniques for generation of Cas9 strains. Transient translocation of the Cas9 protein itself via *Agrobacterium* could be powerful alternative to the RNP method described above. This could be mediated using the following strategy: Firstly, Cas9 CDS is fused to a type-VI secretion signal tag which would then need to be cloned into the domesticated 'helper' plasmid of a suitable *Agrobacterium* strain. The genes of the helper plasmid (which contains the virulence genes) are under the control of a chemical compound such as acetosyringone, which is only added when inducing transformation, avoiding unnecessary expression in the *Agrobacterium*. Secondly, the DNA encoding the single guide RNAs (sgRNAs) is included on the T-DNA portion of the binary Ti plasmid, which gets transferred to the host and integrated randomly into nuclear host DNA during transformation. After the transformation has been carried out, the sgRNAs would be produced and able to be used by the transferred Cas9 proteins, which should degrade over time to avoid toxicity, and importantly no additional Cas9 will be produced once the transformation procedure is over. The sgRNA needs to be designed to knock out an appropriate marker gene to allow for selection, such as to create a photosynthetic deficient or a certain metabolite deficient strain, in order to select for correct clones. The promiscuity of *Agrobacterium* in terms of its infection host could mean that this opens to the door to genome editing technologies in organism currently without the tools to facilitate use of the other methods.

7.4.2 Characterisation of unusual cell metabolites: The example of fatty-acid amides

As explained in the introduction and evident in the lipid investigations in chapter three, microalgae have a diverse lipid metabolism. In the introduction it was highlighted in one survey of the fatty acid profile of a culture collection that there was a large presence of the fatty-acid primary amides in *C. sorokiniana* 211/8k SAG (equivalent to UTEX1230). It would be interesting to understand the function of these species considering they accounted for so much of the percentage in the cell. It also poses the question as to whether these fatty acid amides are falling 'under the radar' in lipid research. For example, they were not identified in the analysis in Chapter 3 which was based on FAME by GC-FID analysis. It could be interesting to explore this further; for example if these primary fatty acid amides are present in large quantities perhaps this could be taken advantage of by increasing expression of a fatty acid amide hydrolase to create a larger free fatty acid pool (Ezzili et al., 2010; McKinney and Cravatt, 2005).

It seems that these fatty-acid amides are not present in algal lipid metabolism schema. This could be because they are generally not present in the model organisms. For example, an examination of the raw data (Lang et al., 2011) shows that out of the 16 strains of *C. reinhardtii*, none contain these two fatty acid primary amides. Another such organism with a detailed metabolic map is *P. tricornutum* and looking at data of three *Phaeodactylum* sp. strains 1090-1a 1090-1b and 1090-6 have 0% or <1% of the oleamide and 0% of palmitamide (Lang et al., 2011). This highlights yet again the problems of the diversity of microalgae and the fact that assumptions made for one species cannot be drawn for another. A brief glance at the data across all the species with the genus name *Chlorella* shows that the oleamide is present in large amounts, with the palmitamide at lower concentrations. Since the data was a high-throughput screen it would certainly be worth further investigation.

It could be worth looking at the new genome of *C. sorokiniana* UTEX1230 (=SAG 211-8k) and seeing if there is an endogenous fatty acid amide hydrolase. If the metabolic enzymes associated with these molecules are present, then it would

further confirm that it is not a data artefact from the high throughput screening methodology.

The implication of such uncharacterised metabolites is not limited to academic studies on metabolic maps. Many algae are sold as food or nutritional supplements and it is interesting to therefore consider the health effect that these fatty-acid amides may have. Both palmitamide and oleamide have been reported as the cause of toxicity in the golden alga *Prymnesium parvum* (Bertin et al., 2012), and oleoamide in particular has a role in significant mammalian process such as induction of sleep. It can also bind the cannabinoid type 1 receptor CB1 in animals (Ezzili et al., 2010).

This emphasises the importance of rigorous safety and toxicity tests of any microalgae to be sold as food for animals or humans and the regulations of 'Generally Regarded As Safe' (GRAS) status. The variation in fatty acid amides between strains of the same species highlights that these classifications need to occur on a strain by strain basis and that species alone is not enough. Additionally, as has been described in the introduction and observed in the results of this thesis, the highly variable metabolism in response to growth conditions could mean that these fatty acid amides remain undetected in certain circumstances, and therefore perhaps safety classifications should include culture conditions as part of their analyses.

Another example of an unexpected effect as described in the introduction to Chapter 4 is the association between the presence of chlorovirus ACTV-1 in throat swabs of humans, and decline in cognitive performance, an observation replicated by directly feeding mice with chlorovirus infected *Chlorella heliozoae*. This could be a further element which needs screening for in safety tests, and emphasises the importance of Good Manufacturing Practice principles in the production and distribution of algal biomass for nutritional purposes.

7.5 Concluding Remarks

In line with the original aims of this thesis, it can be concluded that the genetic engineering of *C. sorokiniana* is hindered by difficulties in transformation and

selection of stable transformants and that production of the foreign fatty-acids EPA and DHA remains a distant prospect using this organism as the host.

In a wider context, it should also be highlighted that engineering of a microalgal strain in a way that perturbs the central carbon metabolism is likely to lead to unexpected effects due to the nature of this responsive complex network. However, in this regard, *C. sorokiniana* UTEX 1230 could theoretically still be a promising host for omega-3 production because the absence of native omega-3 pathways means there may be less opportunity for metabolic interference than in other hosts where there may be other drivers of flux through the pathway.

The development of stable transformation protocols is critical to future progress in engineering microalgae and employing reverse genetic strategies to probe the molecular basis of the metabolism involved in the production of desirable bioproducts. Incorporation of well characterised, predictable genetic parts would help to achieve this goal, and steps towards this were made in this thesis by the creation of a shareable library of parts including new putative regulatory regions from chlorovirus genes. The use of a combinatorial cloning strategy is also valuable for situations in which multi-gene modifications are needed for insertion of whole metabolic pathways.

The work presented in this thesis demonstrates efforts towards creating transgenic lines of *C. sorokiniana* and highlights important experimental issues with using antibiotic selection and using *Agrobacterium*-mediated transformation in this microalga. This work also demonstrated that looking for DNA parts from chloroviruses could be a promising strategy for expanding the genetic toolbox for microalgae. Additonally, if a transgenic strain of *C. sorokiniana* becomes available which can synthesise PUFAs longer than ALA, this thesis identified some suitable growth conditions for such a strain based on FAME profile and productivity, and plenty of avenues for further investigation.

Overall, to produce high-value chemicals in microalgae – depending on the metabolic nature of the desired bioproduct and the state of genetic tools available – rather than genetic engineering it may be worth pursuing strategies such as more

rigorous high-throughput screening of growth conditions. Bioprospecting from culture collections or environmental sources to find a strain with ideal characteristics is also a promising strategy considering the mind-blowing diversity of these interesting and beautiful single-celled photosynthetic creatures we call ‘microalgae’.

REFERENCES

Adarme-Vega, T.C., Lim, D.K.Y., Timmins, M., Vernen, F., Li, Y., Schenk, P.M., 2012. Microalgal biofactories: a promising approach towards sustainable omega-3 fatty acid production. Microb. Cell Fact. 11, 96.

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., Blankenberg, D., 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 46, W537–W544.

Agarkova, I., Hertel, B., Zhang, X., Lane, L., Tchourbanov, A., Dunigan, D.D., Thiel, G., Rossmann, M.G., Van Etten, J.L., 2014. Dynamic attachment of Chlorovirus PBCV-1 to Chlorella variabilis. Virology 466–467, 95–102.

Agarkova, I. V, Dunigan, D.D., Van Etten, J.L., 2006. Virion-associated restriction endonucleases of chloroviruses. J. Virol. 80, 8114–23.

Akbari, F., Eskandani, M., Khosroushahi, A.Y., 2014. The potential of transgenic green microalgae; a robust photobioreactor to produce recombinant therapeutic proteins. World J. Microbiol. Biotechnol. 30, 2783–96.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., 2008. Molecular biology of the cell, Fifth. ed. Garland Science, New York.

Alemán-Nava, G.S., Cuellar-Bermudez, S.P., Cuaresma, M., Bosma, R., Muylaert, K., Ritmann, B.E., Parra, R., 2016. How to us Nile Red, a selective fluorescent stain for microalgal neutral lipids. J. Microbiol. Methods 128, 74–79.

Amminger, G.P., Schafer, M.R., Schlogelhofer, M., Klier, C.M., McGorry, P.D., 2015. Longer-term outcome in the prevention of psychotic disorders by the Vienna omega-3 study. Nat. Commun. 6, 7934.

Andersen, R.A., 2013. The Microalgal Cell. In: Richmond, A., Hu, Q. (Eds.), Handbook of Microalgal Culture: Applied Phycology and Biotechnology. Blackwell Publishing Ltd., pp. 3–20.

Andreou, A.I., Nakayama, N., 2018. Mobius Assembly: A versatile Golden-Gate framework towards universal DNA assembly. PLoS One 13, e0189892.

Anila, N., Chandrashek, A., Ravishankar, G. a, Sarada, R., 2011. Establishment of Agrobacterium tumefaciens-mediated genetic transformation in Dunaliella bardawil. Eur. J. Phycol. 46, 36–44.

Anila, N., Simon, D.P., Chandrashekar, A., Ravishankar, G.A., Sarada, R., 2015. Metabolic engineering of Dunaliella salina for production of ketocarotenoids. Photosynth. Res. 127, 321–333.

Arora, N., Patel, A., Pruthi, P.A., Pruthi, V., 2015. Synergistic dynamics of nitrogen and phosphorous influences lipid productivity in Chlorella minutissima for biodiesel production. Bioresour. Technol. 213, 79–87.

Axelsson, M., Gentili, F., 2014. A single-step method for rapid extraction of total lipids from green microalgae. PLoS One 9, 17–20.

Bai, L.-L., Yin, W.-B., Chen, Y.-H., Niu, L.-L., Sun, Y.-R., Zhao, S.-M., Yang, F.-Q., Wang, R.R.-C., Wu, Q., Zhang, X.-Q., Hu, Z.-M., 2013. A New Strategy to Produce a Defensin: Stable Production of Mutated NP-1 in Nitrate Reductase-Deficient Chlorella ellipsoidea. PLoS One 8, e54966.

Baier, T., Wichmann, J., Kruse, O., Lauersen, K.J., 2018. Intron-containing algal transgenes mediate efficient recombinant gene expression in the green microalga Chlamydomonas reinhardtii. Nucleic Acids Res. 46, 6909–6919.

Baker, M., Penny, D., 2016. Is there a reproducibility crisis? Nature 533, 452–454.

Banerjee, C., Singh, P.K., Shukla, P., 2016. Microalgal bioengineering for sustainable energy development: Recent transgenesis and metabolic engineering strategies. Biotechnol. J. 11, 303–314.

Bashan, Y., Lopez, B.R., Huss, V.A.R., Amavizca, E., de-Bashan, L.E., 2016. Chlorella sorokiniana (formerly C. vulgaris) UTEX 2714, a non-thermotolerant microalga useful for biotechnological applications and as a reference strain. J. Appl. Phycol. 28, 113–121.

Bashir, K.M.I., Kim, M.S., Stahl, U., Cho, M.G., 2016. Microalgae engineering toolbox: Selectable and screenable markers. Biotechnol. Bioprocess Eng. 21, 224–235.

Baudelet, P.H., Ricochon, G., Linder, M., Muniglia, L., 2017. A new insight into cell walls of Chlorophyta. Algal Res. 25, 333–371.

Beal, J., Haddock-Angelli, T., Gershater, M., De Mora, K., Lizarazo, M., Hollenhorst, J., Rettberg, R., 2016. Reproducibility of fluorescent expression from engineered biological constructs in E. coli. PLoS One 11.

Becker, W.E., 2013. Microalgae in Human and Animal Nutrition. In: Richmond, A., Hu, Q. (Eds.), Handbook of Microalgal Culture: Applied Phycology and Biotechnology. Blackwell Publishing Ltd., pp. 461–503.

Beijerinck, M.W., 1890. Culturversuche mit Zoochlorellen, Lichen- engonidien und anderen niederen Algen I–III. Bot. Zeitung 48, 726–740.

Bellou, S., Aggelis, G., 2012. Biochemical activities in Chlorella sp. and Nannochloropsis salina during lipid and sugar synthesis in a lab-scale open pond simulating reactor. J. Biotechnol. 164, 318–29.

Bellou, S., Baeshen, M.N., Elazzazy, A.M., Aggeli, D., Sayegh, F., Aggelis, G., 2014. Microalgal lipids biochemistry and biotechnological perspectives. Biotechnol. Adv. 32, 1476–1493.

Bellou, S., Triantaphyllidou, I.-E., Aggeli, D., Elazzazy, A.M., Baeshen, M.N., Aggelis, G., 2016. Microbial oils as food additives: recent approaches for improving microbial oil production and its polyunsaturated fatty acid content. Curr. Opin. Biotechnol. 37, 24–35.

Bertalan, I., Munder, M.C., Weiß, C., Kopf, J., Fischer, D., Johanningmeier, U., 2015. A rapid, modular and marker-free chloroplast expression system for the green alga Chlamydomonas reinhardtii. J. Biotechnol. 195, 60–66.

Bertin, M.J., Zimba, P. V., Beauchesne, K.R., Huncik, K.M., Moeller, P.D.R., 2012. The contribution of fatty acid amides to Prymnesium parvum Carter toxicity. Harmful Algae 20, 117–125.

Bharathiraja, B., Sridharan, S., Sowmya, V., Yuvaraj, D., Praveenkumar, R., 2017. Microbial oil – A plausible alternate resource for food and fuel application. Bioresour. Technol. 233, 423–432.

Bihmidine, S., Cao, M., Kang, M., Awada, T., Van Etten, J.L., Dunigan, D.D., Clemente, T.E., 2014. Expression of Chlorovirus MT325 aquaglyceroporin (aqpv1) in tobacco and its role in mitigating drought stress. Planta 240, 209–21.

Blanc, G., Duncan, G., Agarkova, I., Borodovsky, M., Gurnon, J., Kuo, A., Lindquist, E., Lucas, S., Pangilinan, J., Polle, J., Salamov, A., Terry, A., Yamada, T., Dunigan, D.D., Grigoriev, I. V., Claverie, J.-M., Van Etten, J.L., 2010. The Chlorella variabilis NC64A Genome Reveals Adaptation to Photosymbiosis, Coevolution with Viruses, and Cryptic Sex. Plant Cell 22, 2943–2955.

Blanc, G., Mozar, M., Agarkova, I. V., Gurnon, J.R., Yanai-Balser, G., Rowe, J.M., Xia, Y., Riethoven, J.J., Dunigan, D.D., Van Etten, J.L., 2014. Deep RNA sequencing reveals hidden features and dynamics of early gene transcription in Paramecium bursaria chlorella virus 1. PLoS One 9, 1–10.

Bonaventura, C., Myers, J., 1969. Fluorescence and oxygen evolution from Chlorella pyrenoidosa. Biochim. Biophys. Acta 189, 366–83.

Borowitzka, M.A., 2013. High-value products from microalgae - their development and commercialisation. J. Appl. Phycol. 25, 743–756.

Boyle, N.R., Morgan, J.A., 2009. Flux balance analysis of primary metabolism in Chlamydomonas reinhardtii. BMC Syst. Biol. 3, 1–14.

Brand, J.J., Andersen, R.A., Nobles, D.R., 2013. Maintenance of Microalgae in Culture Collections. In: Richmond, A., Hu, Q. (Eds.), Handbook of Microalgal Culture: Applied Phycology and Biotechnology. Blackwell Publishing Ltd., pp. 80–89.

Breuer, G., de Jaeger, L., Artus, V.P.G., Martens, D.E., Springer, J., Draaisma, R.B., Eggink, G., Wijffels, R.H., Lamers, P.P., 2014. Superior triacylglycerol (TAG) accumulation in starchless mutants of Scenedesmus obliquus: (II) evaluation of TAG yield and productivity in controlled photobioreactors. Biotechnol. Biofuels 7, 70.

Brunner, E.J., Jones, P.J.S., Friel, S., Bartley, M., 2009. Fish, human health and marine ecosystem health: policies in collision. Int. J. Epidemiol. 38, 93–100.

Bubeck, J.A., Pfitzner, A.J.P., 2005. Isolation and characterization of a new type of chlorovirus that infects an endosymbiotic Chlorella strain of the heliozoon Acanthocystis turfacea. J. Gen. Virol. 86, 2871–2877.

Campbell, R.E., Campbell, R.E., Tour, O., Tour, O., Palmer, A.E., Palmer, A.E., Steinbach, P. a, Steinbach, P. a, Baird, G.S., Baird, G.S., Zacharias, D. a, Zacharias, D. a, Tsien, R.Y., Tsien, R.Y., 2002. A monomeric red fluorescent protein. Proc. Natl. Acad. Sci. U. S. A. 99, 7877–82.

Cavonius, L.R., Carlsson, N.-G., Undeland, I., 2014. Quantification of total fatty acids in microalgae: comparison of extraction and transesterification methods. Anal. Bioanal. Chem. 406, 7313–7322.

Cecchin, M., Benfatto, S., Griggio, F., Mori, A., Cazzaniga, S., Vitulo, N., Delledonne, M., Ballottari, M., 2018. Molecular basis of autotrophic vs mixotrophic growth in Chlorella sorokiniana. Sci. Rep. 8, 1–13.

Cerutti, H., Johnson, A.M., Gillham, N.W., Boynto, J.E., 1997. Epigenetic Silencing of a Foreign Gene in Nuclear Transformants of Chlamydomonas. Plant Cell 9, 925–945.

Cha, T.S., Chen, C.F., Yee, W., Aziz, A., Loh, S.H., 2011a. Cinnamic acid, coumarin and vanillin: Alternative phenolic compounds for efficient Agrobacterium-mediated transformation of the unicellular green alga, Nannochloropsis sp. J. Microbiol. Methods 84, 430–434.

Cha, T.S., Yee, W., Ahmad, A., 2011b. Assessment of Factors Affecting Agrobacterium - Mediated Transformation of Microalgae. Conf. Pap. 10th Int. Symp. Sustain. Sci. Manag. 633–637.

Cha, T.S., Yee, W., Aziz, A., 2012. Assessment of factors affecting Agrobacterium-mediated genetic transformation of the unicellular green alga, Chlorella vulgaris. World J. Microbiol. Biotechnol. 28, 1771–9.

Champenois, J., Marfaing, H., Pierre, R., 2015. Review of the taxonomic revision of Chlorella and consequences for its food uses in Europe. J. Appl. Phycol. 27, 1845–1851.

Charoonnart, P., Purton, S., Saksmerprome, V., 2018. Applications of Microalgal Biotechnology for Disease Control in Aquaculture. Biology (Basel). 7, 24.

Chauton, M.S., Reitan, K.I., Norsker, N.H., Tveterås, R., Kleivdal, H.T., 2015. A techno-economic analysis of industrial production of marine microalgae as a source of EPA and DHA-rich raw material for aquafeed: Research challenges and possibilities. Aquaculture 436, 95–103.

Chen, B., Wan, C., Mehmood, M.A., Chang, J.S., Bai, F., Zhao, X., 2017. Manipulating environmental stresses and stress tolerance of microalgae for enhanced production of lipids and value-added products - A review. Bioresour. Technol. 244, 1198–1206.

Chen, F., Johns, M.R., 1991. Effect of C/N ratio and aeration on the fatty acid composition of heterotrophic Chlorella sorokiniana. J. Appl. Phycol. 3, 203–209.

Chen, H., Qiu, T., Rong, J., He, C., Wang, Q., 2015. Microalgal biofuel revisited: An informatics-based analysis of developments to date and future prospects. Appl. Energy 155, 585–598.

Chen, T., Liu, J., Guo, B., Ma, X., Sun, P., Liu, B., Chen, F., 2015. Light attenuates lipid

accumulation while enhancing cell proliferation and starch synthesis in the glucose-fed oleaginous microalga Chlorella zofingiensis. Sci. Rep. 5, 14936.

Chen, W., Zhang, C., Song, L., Sommerfeld, M., Hu, Q., 2009. A high throughput Nile red method for quantitative measurement of neutral lipids in microalgae. J. Microbiol. Methods 77, 41–47.

Chen, Y., Wang, Y., Sun, Y., Zhang, L., Li, W., 2001. Highly efficient expression of rabbit neutrophil peptide-1 gene in Chlorella ellipsoidea cells. Curr. Genet. 39, 365–70.

Cheng, R., Ma, R., Li, K., Rong, H., Lin, X., Wang, Z., Yang, S., Ma, Y., 2012. Agrobacterium tumefaciens mediated transformation of marine microalgae Schizochytrium. Microbiol. Res. 167, 179–86.

Cheng, Y., Labavitch, J., 2015. Organic and Inorganic Nitrogen Impact Chlorella variabilis Productivity and Host Quality for Viral Production and Cell Lysis. Appl Biochem Biotechnol 176, 467–479.

Chew, K.W., Chia, S.R., Show, P.L., Yap, Y.J., Ling, T.C., Chang, J.S., 2018. Effects of water culture medium, cultivation systems and growth modes for microalgae cultivation: A review. J. Taiwan Inst. Chem. Eng. 0, 1–13.

Chiaiese, P., Palomba, F., Tatino, F., Lanzillo, C., Pinto, G., Pollio, A., Filippone, E., 2011. Engineered tobacco and microalgae secreting the fungal laccase POXA1b reduce phenol content in olive oil mill wastewater. Enzyme Microb. Technol. 49, 540–546.

Chioccioli, M., Hankamer, B., Ross, I.L., 2014. Flow cytometry pulse width data enables rapid and sensitive estimation of biomass dry weight in the microalgae Chlamydomonas reinhardtii and Chlorella vulgaris. PLoS One 9, 1–12.

Chiranjeevi, P., Venkata Mohan, S., 2016. Optimizing the Critical Factors for Lipid Productivity during Stress Phased Heterotrophic Microalgae Cultivation. Front. Energy Res. 4, 1–10.

Cho, H.-H., Park, H.-H., Kim, J.-O., Choi, T.-J., 2002. Isolation and characterization of Chlorella viruses from freshwater sources in Korea. Mol. Cells 14, 168–176.

Chow, K.C., Tung, W.L., 1999. Electrotransformation of Chlorella vulgaris. Plant Cell Rep. 18, 778–780.

Clerck, O. De, Bogaert, K.A., Leliaert, F., 2012. Diversity and evolution of algae: primary endosymbiosis, Advances in Botanical Research. Elsevier Ltd.

Coleman, A.W., 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. Trends Genet. 19, 370–375.

Coll, J.M., 2006. Methodologies for transferring DNA into eukaryotic microalgae: a review. Spanish J. Agric. Res. 4, 316.

Cox, R.S., Madsen, C., McLaughlin, J., Nguyen, T., Roehner, N., Bartley, B., Bhatia, S., Bissell, M., Clancy, K., Gorochowski, T., Grünberg, R., Luna, A., Le Novère, N., Pocock, M., Sauro, H., Sexton, J.T., Stan, G.-B., Tabor, J.J., Voigt, C.A., Zundel, Z., Myers, C., Beal, J., Wipat, A., 2018. Synthetic Biology Open Language Visual (SBOL Visual) Version 2.0. J. Integr. Bioinform. 15.

Cross, F.R., 2016. Tying Down Loose Ends in the Chlamydomonas Genome: Functional Significance of Abundant Upstream Open Reading Frames. G3 Genes, Genomes, Genet. 6, 435–46.

Crozet, P., Navarro, F., Willmund, F., Mehrshahi, P., Bakowski, K., Lauersen, K., Pérez-Pérez, M., Auroy, P., Gorchs Rovira, A., Sauret-Gueto, S., Niemeyer, J., Spaniol, B., Theis, J., Trösch, R., Westrich, L., Vavitsas, K., Baier, T., Hübner, W., de Carpentier, F., Cassarini, M., Danon, A., Henri, J., Marchand, C., de Mia, M., Sarkissian, K., Baulcombe, D., Peltier, G., Crespo, J., Kruse, O., Jensen, P., Schroda, M., Smith, A., Lemaire, S., 2018. Birth of a photosynthetic chassis: a MoClo toolkit enabling synthetic biology in the microalga Chlamydomonas reinhardtii. ACS Synth. Biol. In Press.

Cuaresma, M., Janssen, M., Vílchez, C., Wijffels, R.H., 2009. Productivity of Chlorella sorokiniana in a short light-path (SLP) panel photobioreactor under high irradiance. Biotechnol. Bioeng. 104, 352–9.

D'Alessandro, E.B., Antoniosi Filho, N.R., 2016. Concepts and studies on lipid and pigments of microalgae: A review. Renew. Sustain. Energy Rev. 58, 832–841.

Da Costa, E., Silva, J., Mendonï¿½a, S.H., Abreu, M.H., Domingues, M.R., 2016. Lipidomic approaches towards deciphering glycolipids from microalgae as a reservoir of bioactive lipids. Mar. Drugs 14.

Dalla Costa, L., Piazza, S., Campa, M., Flachowsky, H., Hanke, M.V., Malnoy, M., 2015. Efficient heat-shock removal of the selectable marker gene in genetically modified grapevine. Plant Cell. Tissue Organ Cult. 124, 471–481.

Dalton, R., 2002. Squaring up over ancient life. Nature 417, 782–784.

Dautor, Y., Ubeda-Minguez, P., Chileh, T., Garcia-Maroto, F., Alonso, D.L., 2014. Development of genetic transformation methodologies for an industrially-promising microalga: Scenedesmus almeriensis. Biotechnol. Lett. 36, 2551–2558.

Davis, I.W., Benninger, C., Benfey, P.N., Elich, T., 2012. Powrs: Position-sensitive motif discovery. PLoS One 7, 1–10.

de-Bashan, L.E., Trejo, A., Huss, V.A.R., Hernandez, J.P., Bashan, Y., 2008. Chlorella sorokiniana UTEX 2805, a heat and intense, sunlight-tolerant microalga with potential for removing ammonium from wastewater. Bioresour. Technol. 99, 4980–4989.

De Bhowmick, G., Koduru, L., Sen, R., 2015. Metabolic pathway engineering towards enhancing microalgal lipid biosynthesis for biofuel application—A review. Renew. Sustain. Energy Rev. 50, 1239–1253.

De Castro, C., Molinaro, A., Piacente, F., Gurnon, J.R., Sturiale, L., Palmigiano, A., Lanzetta, R., Parrilli, M., Garozzo, D., Tonetti, M.G., Van Etten, J.L., 2013. Structure of N-linked oligosaccharides attached to chlorovirus PBCV-1 major capsid protein reveals unusual class of complex N-glycans. Proc. Natl. Acad. Sci. U. S. A. 110, 13956–60.

De Castro, C., Speciale, I., Duncan, G., Dunigan, D.D., Agarkova, I., Lanzetta, R., Sturiale, L., Palmigiano, A., Garozzo, D., Molinaro, A., Tonetti, M., Van Etten, J.L., 2015. N-linked Glycans of Chloroviruses Sharing a Core Architecture without Precedent. Angew. Chemie 127.

de Jaeger, L., Verbeek, R.E., Draaisma, R.B., Martens, D.E., Springer, J., Eggink, G., Wijffels, R.H., 2014. Superior triacylglycerol (TAG) accumulation in starchless mutants of Scenedesmus obliquus: (I) mutant generation and characterization. Biotechnol. Biofuels 7, 69.

de Lomana, A.L.G., Schauble, S., Valenzuela, J., Imam, S., Carter, W., Bilgin, D.D., Yohn, C.B., Turkarslan, S., Reiss, D.J., Orellana, M. V., Price, N.D., Baliga, N.S., 2015. Transcriptional program for nitrogen starvation-induced lipid accumulation in Chlamydomonas reinhardtii. Biotechnol. Biofuels 8, 207.

de Moraes, J., de Oliveira, R.N., Costa, J.P., Junior, A.L.G., de Sousa, D.P., Freitas, R.M., Allegretti, S.M., Pinto, P.L.S., 2014. Phytol, a Diterpene Alcohol from Chlorophyll, as a Drug against Neglected Tropical Disease Schistosomiasis Mansoni. PLoS Negl. Trop. Dis. 8, 51.

Deeba, F., Hyder, M.Z., Shah, S.H., Naqvi, S.M.S., 2014. Multiplex PCR assay for identification of commonly used disarmed Agrobacterium tumefaciens strains. Springer Plus 3, 1–7.

Dejoye Tanzi, C., Abert Vian, M., Chemat, F., 2013. New procedure for extraction of algal lipids from wet biomass: A green clean and scalable process. Bioresour. Technol. 134, 271–275.

Delrue, F., Álvarez-Díaz, P.D., Fon-Sing, S., Fleury, G., Sassi, J.F., 2016. The environmental biorefinery: Using microalgae to remediate wastewater, a win-win paradigm. Energies 9, 1–19.

Díaz-Santos, E., de la Vega, M., Vila, M., Vigara, J., León, R., 2013. Efficiency of different heterologous promoters in the unicellular microalga Chlamydomonas reinhardtii. Biotechnol. Prog. 29, 319–328.

Ding, G.T.A.O., Takriff, M.S., Salihon, J., Syukri, M., Rahaman, A.B.D., 2016. Feasiblity of the

Optical Density (OD) in the Determination of the Microalgal Biomass Using Palm Oil Mill Effluent (POME) As Medium. Proc. 50th IIER Int. Conf. 36–38.

Domozych, D.S., Ciancia, M., Fangel, J.U., Mikkelsen, M.D., Ulvskov, P., Willats, W.G.T., 2012. The Cell Walls of Green Algae: A Journey through Evolution and Diversity. Front. Plant Sci. 3, 1–7.

Doron, L., Segal, N., Shapira, M., 2016. Transgene Expression in Microalgae-From Tools to Applications. Front. Plant Sci. 7, 505.

Draaisma, R.B., Wijffels, R.H., Slegers, P.M.E., Brentner, L.B., Roy, A., Barbosa, M.J., 2013. Food commodities from microalgae. Curr. Opin. Biotechnol. 24, 169–77.

Dunigan, D.D., Cerny, R.L., Bauman, A.T., Roach, J.C., Lane, L.C., Agarkova, I. V., Wulser, K., Yanai-Balser, G.M., Gurnon, J.R., Vitek, J.C., Kronschnabel, B.J., Jeanniard, A., Blanc, G., Upton, C., Duncan, G.A., McClung, O.W., Ma, F., Van Etten, J.L., 2012. Paramecium bursaria Chlorella Virus 1 Proteome Reveals Novel Architectural and Regulatory Features of a Giant Virus. J. Virol. 86, 8821–8834.

El-Sheekh, M.M., Gheda, S.F., Khairy, H.M., El-Shenody, R.A., 2015. Optimization of Medium Components Using Plackett-Burman Design for High Production of Protein, Carbohydrates and Lipids in the Microalga Tetraselmis Chuii. Egypt. J. Exp. Biol. 11, 77–88.

Engler, C., Gruetzner, R., Kandzia, R., Marillonnet, S., 2009. Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes. PLoS One 4.

Engler, C., Youles, M., Gruetzner, R., Ehnert, T.M., Werner, S., Jones, J.D.G., Patron, N.J., Marillonnet, S., 2014. A Golden Gate modular cloning toolbox for plants. ACS Synth. Biol. 3, 839–843.

Esland, L., Larrea-Alvarez, M., Purton, S., 2018. Selectable markers and reporter genes for engineering the chloroplast of Chlamydomonas. Biology (Basel). 6, In press.

European Commission, 2015. Novel Food Catalogue [WWW Document]. URL http://ec.europa.eu/food/safety/novel_food/catalogue/search/public/index.cfm# (accessed 8.28.18).

Evens, T.J., Niedz, R.P., 2010. Quantification of nutrient-replete growth rates in five-ion hyperspace for Chlorella vulgaris (Trebouxiophyceae) and Peridinium cinctum (Dinophyceae). Eur. J. Phycol. 45, 247–257.

Ezzili, C., Otrubova, K., Boger, D.L., 2010. Fatty acid amide signaling molecules. Bioorganic Med. Chem. Lett. 20, 5959–5968.

Falkowski, P.G., Katz, M.E., Knoll, A.H., Quigg, A., Raven, J.A., Schofield, O., Taylor, F.J.R., 2004. The Evolution of Modern Eukaryotic Phytoplankton. Science (80-.). 305, 354–360.

Fan, J., Cui, Y., Wan, M., Wang, W., Li, Y., 2014. Lipid accumulation and biosynthesis genes response of the oleaginous Chlorella pyrenoidosa under three nutrition stressors. Biotechnolgy for Biofuels 7, 17.

Fan, J., Ning, K., Zeng, X., Luo, Y., Wang, D., Hu, J., Li, J., Xu, H., Huang, J., Wan, M., Wang, W., Zhang, D., Shen, G., Run, C., Liao, J., Fang, L., Huang, S., Jing, X., Su, X., Wang, A., Bai, L., Hu, Z.M., Xu, J., Li, Y., 2015. Genomic Foundation of Starch to Lipid Switch in Oleaginous Chlorella. Plant Physiol. pp.01174.2015.

Fang, L., Lin, H.X., Low, C.S., Wu, M.H., Chow, Y., Lee, Y.K., 2012. Expression of the Chlamydomonas reinhardtii Sedoheptulose-1,7-bisphosphatase in Dunaliella bardawil leads to enhanced photosynthesis and increased glycerol production. Plant Biotechnol. J. 10, 1129–1135.

Ferenczi, A., Pyott, D.E., Xipnitou, A., Molnar, A., 2017. Efficient targeted DNA editing and replacement in *Chlamydomonas reinhardtii* using Cpf1 ribonucleoproteins and single-stranded DNA. Proc. Natl. Acad. Sci. 114, 201710597.

Field, C.B., Behrenfeld, M.J., Randerson, J.T., Falkowski, P., 1998. Primary production of the biosphere: Integrating terrestrial and oceanic components. Science (80-.). 281, 237–240.

Fields, F.J., Kociolek, J.P., 2015. An evolutionary perspective on selecting high-lipid-content diatoms (Bacillariophyta). J. Appl. Phycol. 27, 2209–2220.

Fisher, A.W.J., Burlew, J.S., 1953. Nutritional Value of Microscopic Algae. In: Burlew, J.S. (Ed.), Algal Culture: From Laboratory to Pilot Plant. CARNEGIE INSTITUTION OF WASHINGTON PUBLICATION 600, Washington, DC, pp. 303–311.

Fitzgerald, L.A., Boucher, P.T., Yanai-Balser, G.M., Suhre, K., Graves, M. V., Van Etten, J.L., 2008. Putative gene promoter sequences in the chlorella viruses. Virology 380, 388–93.

Forján, E., Navarro, F., Cuaresma, M., Vaquero, I., Ruíz-Domínguez, M.C., Gojkovic, Ž., Vázquez, M., Márquez, M., Mogedas, B., Bermejo, E., Girlich, S., Domínguez, M.J., Vílchez, C., Vega, J.M., Garbayo, I., 2015. Microalgae: Fast-growth sustainable green factories. Crit. Rev. Environ. Sci. Technol. 45, 1705–1755.

Fujiwara, T., Oda, K., Yokota, S., Takatsuki, a., Ikehara, Y., 1988. Brefeldin A causes disassembly of the Golgi complex and accumulation of secretory proteins in the endoplasmic reticulum. J. Biol. Chem. 263, 18545–18552.

Gaj, T., Gersbach, C.A., Barbas, C.F., 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. 31, 397–405.

Galarza, J.I., Delgado, N., Henríquez, V., 2016. Cisgenesis and intragenesis in microalgae: promising advancements towards sustainable metabolites production. Appl. Microbiol. Biotechnol. 100, 10225–10235.

Gangl, D., Zedler, J.A.Z., Rajakumar, P.D., Martinez, E.M.R., Riseley, A., Włodarczyk, A., Purton, S., Sakuragi, Y., Howe, C.J., Jensen, P.E., Robinson, C., 2015. Biotechnological exploitation of microalgae. J. Exp. Bot. Advance Ac.

Gao, C., Wang, Y., Shen, Y., Yan, D., He, X., Dai, J., Wu, Q., 2014. Oil accumulation mechanisms of the oleaginous microalga Chlorella protothecoides revealed through its genome, transcriptomes, and proteomes. BMC Genomics 15, 582.

Garay, L.A., Boundy-Mills, K.L., German, J.B., 2014. Accumulation of high-value lipids in single-cell microorganisms: A mechanistic approach and future perspectives. J. Agric. Food Chem. 62, 2709–2727.

García, J.L., de Vicente, M., Galán, B., 2017. Microalgae, old sustainable food and fashion nutraceuticals. Microb. Biotechnol. 10, 1017–1024.

Gargouri, M., Park, J.-J., Holguin, F.O., Kim, M.-J., Wang, H., Deshpande, R.R., Shachar-Hill, Y., Hicks, L.M., Gang, D.R., 2015. Identification of regulatory network hubs that control lipid metabolism in Chlamydomonas reinhardtii. J. Exp. Bot. erv217-.

Geissmann, Q., 2013. OpenCFU, a New Free and Open-Source Software to Count Cell Colonies and Other Circular Objects. PLoS One 8, 1–10.

Gelvin, S.B., 2003. Agrobacterium-Mediated Plant Transformation: the Biology behind the "Gene-Jockeying" Tool. Microbiol. Mol. Biol. Rev. 67, 16–37.

Gelvin, S.B., 2006. Agrobacterium Virulence Gene Induction. In: Wang, K. (Ed.), Agrobacterium Protocols Volume 1. Humana Press Inc., Totowa, New Jersey, pp. 77–84.

Gimpel, J.A., Henríquez, V., Mayfield, S.P., 2015. Metabolic Engineering of Eukaryotic Microalgae: Potential and Challenges Come with Great Diversity. Front. Microbiol. 6.

Gomma, A.E., Lee, S., Sun, S.M., Yang, S.H., Chung, G., 2015. Development of stable marker-free nuclear transformation strategy in the green microalga Chlorella vulgaris. African J. Biotechnol. 14, 2715–2723.

Goncalves, E.C., Koh, J., Zhu, N., Yoo, M.J., Chen, S., Matsuo, T., Johnson, J. V., Rathinasabapathi, B., 2016. Nitrogen starvation-induced accumulation of triacylglycerol in the green algae: Evidence for a role for ROC40, a transcription factor involved in circadian rhythm. Plant J. 85, 743–757.

Gorman, D.S., Levine, R.P., 1965. Cytochrome f and Plastocyanin : Their Sequence in the Photosynthetic Electron Transport Chain of Chlamydomonas reinhardi. Proceeedings Natl. Acad. Sci. United States Am. 54, 1665–1669.

Govindarajulu, M., Elmore, J.M., Fester, T., Taylor, C.G., 2008. Evaluation of Constitutive Viral

Promoters in Transgenic Soybean Roots and Nodules. Mol. Plant-Microbe Interact. 21, 1027–1035.

Graves, M. V, Meints, R.H., 1992. Characterization of the gene for the most abundant in vitro translation product from a virus infected chlorella-like alga. Gene 113, 149–155.

Greenspan, P., Fowler, S.D., 1985. Spectrofluorometric studies of the lipid probe, nile red. J. Lipid Res. 26, 781–789.

Grishagin, I. V., 2015. Automatic cell counting with ImageJ. Anal. Biochem. 473, 63–65.

Grudzinski, W., Krzeminska, I., Luchowski, R., Nosalewicz, A., Gruszecki, W.I., 2016. Strong-light-induced yellowing of green microalgae Chlorella: A study on molecular mechanisms of the acclimation response. Algal Res. 16, 245–254.

Guarnieri, M.T., Nag, A., Smolinski, S.L., Darzins, A., Seibert, M., Pienkos, P.T., 2011. Examination of Triacylglycerol Biosynthetic Pathways via De Novo Transcriptomic and Proteomic Analyses in an Unsequenced Microalga. PLoS One 6, e25851.

Guarnieri, M.T., Nag, A., Yang, S., Pienkos, P.T., 2013. Proteomic analysis of Chlorella vulgaris: Potential targets for enhanced lipid accumulation. J. Proteomics 93, 245–253.

Guarnieri, M.T., Pienkos, P.T., 2015. Algal omics: unlocking bioproduct diversity in algae cell factories. Photosynth. Res. 123, 255–263.

Guckert, J.B., Cooksey, K.E., 1990. TRIGLYCERIDE ACCUMULATION AND FATTY ACID PROFILE CHANGES IN CHLORELLA (CHLOROPHYTA) DURING HIGH pH-INDUCED CELL CYCLE INHIBITION1. J. Phycol. 26, 72–79.

Halim, R., Webley, P.A., 2015. Nile Red Staining for Oil Determination in Microalgal Cells: A New Insight through Statistical Modelling. Int. J. Chem. Eng. 2015.

Hall, J.D., Fucikova, K., Lo, C., Lewis, L.A., Karol, K.G., 2010. An assessment of proposed DNA barcodes in freshwater green algae. Cryptogam. Algol. 31, 529–555.

Hallenbeck, P.C., Grogger, M., Mraz, M., Veverka, D., 2015a. Building a better Mousetrap I: Using Design of Experiments with unconfounded ions to discover superior media for growth and lipid production by Chlorella sp. EN1234. Bioresour. Technol. 184, 82–89.

Hallenbeck, P.C., Grogger, M., Mraz, M., Veverka, D., 2015b. Building a better mousetrap II: Using Design of Experiments with unconfounded ions to compare the growth of different microalgae. Bioresour. Technol. 184, 90–99.

Hallenbeck, P.C., Grogger, M., Mraz, M., Veverka, D., 2015c. The use of Design of Experiments and Response Surface Methodology to optimize biomass and lipid production by the oleaginous marine green alga, Nannochloropsis gaditana in response to light intensity, inoculum size and CO2. Bioresour. Technol. 184, 161–168.

Hallmann, A., 2007. Algal Transgenics and Biotechnology. Transgenic Plant J. 1, 81–98.

Hamed, I., 2016. The Evolution and Versatility of Microalgal Biotechnology: A Review. Compr. Rev. Food Sci. Food Saf. 15, 1104–1123.

Hamilton, M.L., Haslam, R.P., Napier, J.A., Sayanova, O., 2014. Metabolic engineering of Phaeodactylum tricornutum for the enhanced accumulation of omega-3 long chain polyunsaturated fatty acids. Metab. Eng. 22, 3–9.

Hassall, K.A., 1958. Xylose as a Specific Inhibitor of Photosynthesis. Nature 181, 1273–1274.

Hawkins, R.L., Nakamura, M., 1999. Expression of human growth hormone by the eukaryotic alga, Chlorella. Curr. Microbiol. 38, 335–341.

Heeg, J.S., Wolf, M., 2015. ITS2 and 18S rDNA sequence-structure phylogeny of Chlorella and allies (Chlorophyta, Trebouxiophyceae, Chlorellaceae). Plant Gene 4, 20–28.

Heitzer, M., Eckert, A., Fuhrmann, M., Griesbeck, C., 2007. Influence of Codon Bias on the Expression of Foreign Genes in Microalgae. In: Leon, R., Galvan, A., Fernandez, E. (Eds.), Transgenic Microalgae as Green Cell Factories. Landes Bioscience and Springer Science+Business Media, pp. 46–53.

Hema, R., Senthil-Kumar, M., Shivakumar, S., Chandrasekhara Reddy, P., Udayakumar, M., 2007. Chlamydomonas reinhardtii, a model system for functional validation of abiotic stress responsive genes. Planta 226, 655–670.

Henchion, M., Hayes, M., Mullen, A., Fenelon, M., Tiwari, B., 2017. Future Protein Supply and Demand: Strategies and Factors Influencing a Sustainable Equilibrium. Foods 6, 53.

Henry, E.C., Meints, R.H., 1994. Recombinant viruses as transofrmation vectors of marine macroalgae. J. Appl. Phycol. 6, 247–253.

Heoy-Kung, J., Kim, G., Choi, T., 2006. Activity of Early Gene Promoters from a Korean Chlorella Virus Isolate in Transformed Chlorella Algae. J. Microbiol. Biotechnol. 16, 952–960.

Hernandez-Garcia, C.M., Finer, J.J., 2014. Identification and validation of promoters and cis-acting regulatory elements. Plant Sci. 217–218, 109–119.

Hernández-Torres, A., Zapata-Morales, A.L., Ochoa Alfaro, A.E., Soria-Guerra, R.E., 2016. Identification of gene transcripts involved in lipid biosynthesis in Chlamydomonas reinhardtii under nitrogen, iron and sulfur deprivation. World J. Microbiol. Biotechnol. 32, 1–10.

Hiegle, N., 2014. Development of Tools for Genetic Manipulation of Industrially Relevant Algae. University College London.

Hines, W.C., Su, Y., Kuhn, I., Polyak, K., Bissell, M.J., 2014. Sorting out the FACS: A devil in the details. Cell Rep. 6, 779–781.

Hlavova, M., Turoczy, Z., Bisova, K., 2015. Improving microalgae for biotechnology — From genetics to synthetic biology. Biotechnol. Adv. 33, 1194–1203.

Hsieh, H.-J., Su, C.-H., Chien, L.-J., 2012. Accumulation of lipid production in Chlorella minutissima by triacylglycerol biosynthesis-related genes cloned from Saccharomyces cerevisiae and Yarrowia lipolytica. J. Microbiol. 50, 526–34.

Hu, Q., Sommerfeld, M., Jarvis, E., Ghirardi, M., Posewitz, M., Seibert, M., Darzins, A., 2008. Microalgal triacylglycerols as feedstocks for biofuel production: Perspectives and advances. Plant J. 54, 621–639.

Huang, C.-C., Chen, M.-W., Hsieh, J.-L., Lin, W.-H., Chen, P.-C., Chien, L.-F., 2006. Expression of mercuric reductase from Bacillus megaterium MB1 in eukaryotic microalga Chlorella sp. DT: an approach for mercury phytoremediation. Appl. Microbiol. Biotechnol. 72, 197–205.

Huelsman, T.P., 2015. Genome-Scale Reconstruction of the Chlorella vulgaris UTEX 395 Metabolic Network. University of California, San Diego.

Ikeda, K., Miyasaka, H., 1998. Compilation of mRNA sequences surrounding the AUG translation initiation codon in the green alga Chlamydomonas reinhardtii. Biosci. Biotechnol. Biochem. 62, 2457–2459.

Ikeda, T., Takeda, H., 1995. Species-specific differences of pyrenoids in Chlorella (Chlorophyta). J. Phycol. 31, 813–818.

Illman, A.M., Scragg, A.H., Shales, S.W., 2000. Increase in Chlorella strains calorific values when grown in low nitrogen medium. Enzyme Microb. Technol. 27, 631–635.

Iverson, S.J., Lang, S.L., Cooper, M.H., 2001. Comparison of the Bligh and Dyer and Folch methods for total lipid determination in a broad range of marine tissue. Lipids 36, 1283–1287.

Jeanniard, A., Dunigan, D.D., Gurnon, J.R., Agarkova, I. V., Kang, M., Vitek, J., Duncan, G., McClung, O.W., Larsen, M., Claverie, J.-M., Van Etten, J.L., Blanc, G., 2013. Towards defining the chloroviruses: a genomic journey through a genus of large DNA viruses. BMC Genomics 14, 158.

Jiang, W., Brueggeman, A.J., Horken, K.M., Plucinak, T.M., Weeks, D.P., 2014. Successful transient expression of Cas9 and single guide RNA genes in Chlamydomonas reinhardtii. Eukaryot. Cell 13, 1465–1469.

Johnson, X., Alric, J., 2013. Central Carbon Metabolism and Electron Transport in Chlamydomonas reinhardtii: Metabolic Constraints for Carbon Partitioning between Oil and Starch. Eukaryot. Cell 12, 776–793.

Joshi, C.P., Zhou, H., Huang, X., Chiang, V.L., 1997. Context sequences of translation initiation

codon in plants. Plant Mol. Biol. 35, 993–1001.
Juneja, A., Chaplen, F.W.R., Murthy, G.S., 2016. Genome Scale Metabolic Reconstruction of Chlorella variabilis for exploring its metabolic potential for biofuels. Bioresour. Technol.
Juven-Gershon, T., Cheng, S., Kadonaga, J.T., 2006. Rational design of a super core promoter that enhances gene expression. Nat. Methods 3, 917–922.
K. Y. Lim, D., M. Schenk, P., 2017. Microalgae selection and improvement as oil crops: GM vs non-GM strain engineering. AIMS Bioeng. 4, 151–161.
Kado, C.I., 2014. Historical account on gaining insights on the mechanism of crown gall tumorigenesis induced by Agrobacterium tumefaciens. Front Microbiol 5, 340.
Kadono, T., Miyagawa-Yamaguchi, A., Kira, N., Tomaru, Y., Okami, T., Yoshimatsu, T., Hou, L., Ohama, T., Fukunaga, K., Okauchi, M., Yamaguchi, H., Ohnishi, K., Falciatore, A., Adachi, M., 2015. Characterization of marine diatom-infecting virus promoters in the model diatom Phaeodactylum tricornutum. Sci. Rep. 5, 1–13.
Kanaga, K., Pandey, A., Kumar, S., Geetanjali, 2016. Multi-objective optimization of media nutrients for enhanced production of algae biomass and fatty acid biosynthesis from Chlorella pyrenoidosa NCIM 2738. Bioresour. Technol. 200, 940–950.
Kang, M., Duncan, G.A., Kuszynski, C., Oyler, G., Zheng, J., Becker, D.F., Van Etten, J.L., 2014. Chlorovirus PBCV-1 encodes an active copper-zinc superoxide dismutase. J. Virol. 88, 12541–50.
Kang, M., Graves, M., Mehmel, M., Moroni, A., Gazzarrini, S., Thiel, G., Gurnon, J.R., Van Etten, J.L., 2004. Genetic diversity in chlorella viruses flanking kcv, a gene that encodes a potassium ion channel protein. Virology 326, 150–9.
Kang, M., Han, J.G., Liu, P.F., Ye, Y., Tien, P., 2000. [The regulation activity of Chlorella virus gene 5' upstream sequence in Escherichia coli and eucaryotic alage]. Sheng Wu Gong Cheng Xue Bao 16, 443–6.
Karas, B.J., Diner, R.E., Lefebvre, S.C., McQuaid, J., Phillips, A.P.R., Noddings, C.M., Brunson, J.K., Valas, R.E., Deerinck, T.J., Jablanovic, J., Gillard, J.T.F., Beeri, K., Ellisman, M.H., Glass, J.I., Hutchison III, C.A., Smith, H.O., Venter, J.C., Allen, A.E., Dupont, C.L., Weyman, P.D., 2015. Designer diatom episomes delivered by bacterial conjugation. Nat. Commun. 6, 6925.
Karupaiah, T., Sundram, K., 2007. Effects of stereospecific positioning of fatty acids in triacylglycerol structures in native and randomized fats: A review of their nutritional implications. Nutr. Metab. 4, 1–17.
Kasai, Y., Oshima, K., Ikeda, F., Abe, J., Yoshimitsu, Y., Harayama, S., 2015. Construction of a self-cloning system in the unicellular green alga Pseudochoricystis ellipsoidea. Biotechnol. Biofuels 8, 1–12.
Kathiresan, S., Chandrashekar, A., Ravishankar, G.A., Sarada, R., 2009. Agrobacterium-mediated transformation in the green alga haematococcus pluvialis (chlorophyceae, volvocales). J. Phycol. 45, 642–649.
Kathiresan, S., Chandrashekar, A., Ravishankar, G.A., Sarada, R., 2015. Regulation of astaxanthin and its intermediates through cloning and genetic transformation of β-carotene ketolase in Haematococcus pluvialis. J. Biotechnol. 196–197, 33–41.
Kathiresan, S., Sarada, R., 2009. Towards genetic improvement of commercially important microalga Haematococcus pluvialis for biotech applications. J. Appl. Phycol. 21, 553–558.
Kawasaki, T., Tanaka, M., Fujie, M., Usami, S., Yamada, T., 2004. Immediate early genes expressed in chlorovirus infections. Virology 318, 214–23.
Kay, R., Chan, A., Daly, M., McPherson, J., 1987. Duplication of CaMV 35S Promoter Sequences Creates a Strong Enhancer for Plant Genes. Science (80-.). 236, 1299–1302.
Kern, J.D., Hise, A.M., Characklis, G.W., Gerlach, R., Viamajala, S., Gardner, R.D., 2017. Using life cycle assessment and techno-economic analysis in a real options framework to inform the design of algal biofuel production facilities. Bioresour. Technol. 225, 418–

428.
Kessler, E., 1976. Comparative physiology, biochemistry, and the taxonomy of Chlorella (Chlorophyceae). Plant Syst. Evol. 125, 129–138.
Khozin-Goldberg, I., 2016. Lipid Metabolism in Microalgae. In: Borowitzka, M.A. (Ed.), The Physiology of Microalgae. Springer International Publishing, pp. 413–484.
Khozin-Goldberg, I., Didi-Cohen, S., Shayakhmetova, I., Cohen, Z., 2002. Biosynthesis of eicosapentaenoic acid (EPA) in the freshwater eustigmatophyte Monodus subterraneus (Eustigmatophyceae). J. Phycol. 38, 745–756.
Kim, D.Y., Vijayan, D., Praveenkumar, R., Han, J.I., Lee, K., Park, J.Y., Chang, W.S., Lee, J.S., Oh, Y.K., 2016. Cell-wall disruption and lipid/astaxanthin extraction from microalgae: Chlorella and Haematococcus. Bioresour. Technol. 199, 300–310.
Kim, S.C., Chapman, K.D., Blancaflor, E.B., 2010. Fatty acid amide lipid mediators in plants. Plant Sci. 178, 411–419.
Kindle, K.L., 1998. High-frequency nuclear transformation of Chlamydomonas reinhardtii. Methods Enzymol. 297, 27–38.
Kiran, B., Kumar, R., Deshmukh, D., 2014. Perspectives of microalgal biofuels as a renewable source of energy. Energy Convers. Manag. 88, 1228–1244.
Kirchner, L., Wirshing, A., Kurt, L., Reinard, T., Glick, J., Cram, E.J., Jacobsen, H.J., Lee-Parsons, C.W.T., 2016. Identification, characterization, and expression of diacylgylcerol acyltransferase type-1 from Chlorella vulgaris. Algal Res. 13, 167–181.
Kjartansdóttir, K.R., Friis-Nielsen, J., Asplund, M., Mollerup, S., Mourier, T., Jensen, R.H., Hansen, T.A., Rey-Iglesia, A., Richter, S.R., Alquezar-Planas, D.E., Olsen, P.V.S., Vinner, L., Fridholm, H., Sicheritz-Pontén, T., Nielsen, L.P., Brunak, S., Willerslev, E., Izarzugaza, J.M.G., Hansen, A.J., 2015. Traces of ATCV-1 associated with laboratory component contamination. Proc. Natl. Acad. Sci. U. S. A. 112, E925–E926.
Klein-Marcuschamer, D., Chisti, Y., Benemann, J.R., Lewis, D., 2013. A matter of detail: Assessing the true potential of microalgal biofuels. Biotechnol. Bioeng. 110, 2317–2322.
Klok, A.J., Lamers, P.P., Martens, D.E., Draaisma, R.B., Wijffels, R.H., 2014. Edible oils from microalgae: insights in TAG accumulation. Trends Biotechnol. 32, 521–528.
Kobayashi, N., Noel, E.A., Barnes, A., Watson, A., Rosenberg, J.N., Erickson, G., Oyler, G.A., 2013. Characterization of three Chlorella sorokiniana strains in anaerobic digested effluent from cattle manure. Bioresour. Technol. 150, 377–86.
Komor, E., Tanner, W., 1974. The Hexose-Proton Cotransport System of Chlorella. pH-dependent Change in Km Values and Translocation Constants of the uptake system. J. Gen. Physiol. 64, 568–581.
Koo, J., Park, D., Kim, H., 2013. Expression of bovine lactoferrin N-lobe by the green alga , Chlorella vulgaris. ALGAE 28, 379–387.
Kozak, M., 1987. An analysis of 5’-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. 15, 8125–8148.
Kozak, M., 1989. The scanning model for translation: an update. J. Cell Biol. 108, 229–41.
Krienitz, L., Huss, V.A.R., Bock, C., 2015. Chlorella: 125 years of the green survivalist. Trends Plant Sci. 20, 67–69.
Kropat, J., Hong-hermesdorf, A., Casero, D., Ent, P., Castruita, M., Pellegrini, M., Merchant, S.S., Malasarn, D., 2011. A revised mineral nutrient supplement increases biomass and growth rate in Chlamydomonas reinhardtii. Plant J. 66, 770–780.
Kubota, T., Katou, Y., Nakato, R., Shirahige, K., Donaldson, A.D., 2015. Replication-Coupled PCNA Unloading by the Elg1 Complex Occurs Genome-wide and Requires Okazaki Fragment Ligation. Cell Rep. 12, 774–787.
Kumar, K., Banerjee, D., Das, D., 2014. Carbon dioxide sequestration from industrial flue gas by Chlorella sorokiniana. Bioresour. Technol. 152, 225–33.
Kumar, S.V., Misquitta, R.W., Reddy, V.S., Rao, B.J., Rajam, M.V., 2004. Genetic transformation of the green alga—Chlamydomonas reinhardtii by Agrobacterium

tumefaciens. Plant Sci. 166, 731–738.
Kumar, S.V., Rajam, M. V., 2005. Polyamines enhance Agrobacterium tumefaciens vir gene induction and T-DNA transfer. Plant Sci. 168, 475–480.
Kumar, V.S., Rajam, M. V., 2007. Induction of Agrobacterium tumefaciens vir genes by the green alga, Chlamydomonas reinhardtii. Curr. Sci. 92, 1727–1729.
Kumari, S., Ware, D., 2013. Genome-wide computational prediction and analysis of core promoter elements across plant monocots and dicots. PLoS One 8, e79011.
La Russa, M., Bogen, C., Uhmeyer, A., Doebbe, A., Filippone, E., Kruse, O., Mussgnug, J.H., 2012. Functional analysis of three type-2 DGAT homologue genes for triacylglycerol production in the green microalga Chlamydomonas reinhardtii. J. Biotechnol. 162, 13–20.
Lammers, P.J., Huesemann, M., Boeing, W., Anderson, D.B., Arnold, R.G., Bai, X., Bhole, M., Brhanavan, Y., Brown, L., Brown, J., Brown, J.K., Chisholm, S., Meghan Downes, C., Fulbright, S., Ge, Y., Holladay, J.E., Ketheesan, B., Khopkar, A., Koushik, A., Laur, P., Marrone, B.L., Mott, J.B., Nirmalakhandan, N., Ogden, K.L., Parsons, R.L., Polle, J., Ryan, R.D., Samocha, T., Sayre, R.T., Seger, M., Selvaratnam, T., Sui, R., Thomasson, A., Unc, A., Van Voorhies, W., Waller, P., Yao, Y., Olivares, J.A., 2017. Review of the cultivation program within the National Alliance for Advanced Biofuels and Bioproducts. Algal Res. 22, 166–186.
Lang, I., Hodac, L., Friedl, T., Feussner, I., 2011. Fatty acid profiles and their distribution patterns in microalgae: a comprehensive analysis of more than 2000 strains from the SAG culture collection. BMC Plant Biol. 11, 124.
LANL, 2014. Los Alamos National Laboratory: Greenhouse, An omics knowledge base for algal biofuel feedstocks [WWW Document]. Los Alamos Natl. Lab. URL https://greenhouse.lanl.gov/organisms/ (accessed 8.26.18).
Lau, C.C., Loh, S.H., Aziz, A., Cha, T.S., 2017. Effects of disrupted omega-3 desaturase gene construct on fatty acid composition and expression of four fatty acid biosynthetic genes in transgenic Chlorella vulgaris. Algal Res. 26, 143–152.
Lauersen, K.J., Kruse, O., Mussgnug, J.H., 2015. Targeted expression of nuclear transgenes in Chlamydomonas reinhardtii with a versatile, modular vector toolkit. Appl. Microbiol. Biotechnol. 99, 3491–503.
Laurens, L.M.L., Dempster, T.A., Jones, H.D.T., Wolfrum, E.J., Van Wychen, S., McAllister, J.S.P., Rencenberger, M., Parchert, K.J., Gloe, L.M., 2012. Algal biomass constituent analysis: Method uncertainties and investigation of the underlying measuring chemistries. Anal. Chem. 84, 1879–1887.
Laurens, L.M.L., Van Wychen, S., Pienkos, P.T., Harmon, V.L., McGowen, J., 2017. Harmonization of experimental approach and data collection to streamline analysis of biomass composition from algae in an inter-laboratory setting. Algal Res. 25, 549–557.
Leclercq, J., Szabolcs, T., Martin, F., Montoro, P., 2015. Development of a new pCAMBIA binary vector using Gateway® technology. Plasmid 81, 50–54.
Lee, A.K., Lewis, D.M., Ashman, P.J., 2012. Disruption of microalgal cells for the extraction of lipids for biofuels: Processes and specific energy requirements. Biomass and Bioenergy 46, 89–101.
Lee, Y.-K., Ding, S.-Y., Hoe, C.-H., Low, C.-S., 1996. Mixotrophic growth of Chlorella sorokiniana in outdoor enclosed photobioreactor. J. Appl. Phycol. 8, 163–169.
Leite, G.B., Paranjape, K., Hallenbeck, P.C., 2016. Breakfast of champions: Fast lipid accumulation by cultures of Chlorella and Scenedesmus induced by xylose. Algal Res. 16, 338–348.
Leliaert, F., Smith, D.R., Moreau, H., Herron, M.D., Verbruggen, H., Delwiche, C.F., De Clerck, O., 2012. Phylogeny and Molecular Evolution of the Green Algae. CRC. Crit. Rev. Plant Sci. 31, 1–46.
Leliaert, F., Verbruggen, H., Vanormelingen, P., Steen, F., López-Bautista, J.M., Zuccarello,

G.C., De Clerck, O., 2014. DNA-based species delimitation in algae. Eur. J. Phycol. 49, 179–196.

Leliaert, F., Verbruggen, H., Zechman, F.W., 2011. Into the deep: New discoveries at the base of the green plant phylogeny. BioEssays 33, 683–692.

Lenka, S.K., Carbonaro, N., Park, R., Miller, S.M., Thorpe, I., Li, Y., 2016. Current advances in molecular, biochemical, and computational modeling analysis of microalgal triacylglycerol biosynthesis. Biotechnol. Adv. 34, 1046–1063.

León-Bañares, R., González-Ballester, D., Galván, A., Fernández, E., 2004. Transgenic microalgae as green cell-factories. Trends Biotechnol. 22, 45–52.

Letunic, I., Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44, W242–W245.

Leu, S., Boussiba, S., 2014. Advances in the Production of High-Value Products by Microalgae. Ind. Biotechnol. 10, 169–183.

Levering, J., Broddrick, J., Zengler, K., 2015. Engineering of oleaginous organisms for lipid production. Curr. Opin. Biotechnol. 36, 32–39.

Leyva, L.A., Bashan, Y., De-Bashan, L.E., 2014. Activity of acetyl-CoA carboxylase is not directly linked to accumulation of lipids when Chlorella vulgaris is co-immobilised with Azospirillum brasilense in alginate under autotrophic and heterotrophic conditions. Ann. Microbiol. 339–349.

Li-Beisson, Y., Beisson, F., Riekhof, W., 2015. Metabolism of acyl-lipids in Chlamydomonas reinhardtii. Plant J. 82, 504–522.

Li, F., Owen, R., Simakova, E., 2015. Framing responsible innovation in synthetic biology: the need for a critical discourse analysis approach. J. Responsible Innov. 2, 104–108.

Li, F., Qin, S., Jiang, P., Wu, Y., Zhang, W., 2009. The integrative expression of GUS gene driven by FCP promoter in the seaweed Laminaria japonica (Phaeophyta). J. Appl. Phycol. 21, 287–293.

Li, T., Xu, J., Gao, B., Xiang, W., Li, A., Zhang, C., 2016. Morphology, growth, biochemical composition and photosynthetic performance of Chlorella vulgaris (Trebouxiophyceae) under low and high nitrogen supplies. Algal Res. 16, 481–491.

Li, T., Zheng, Y., Yu, L., Chen, S., 2013. High productivity cultivation of a heat-resistant microalga Chlorella sorokiniana for biofuel production. Bioresour. Technol. 131, 60–7.

Li, T., Zheng, Y., Yu, L., Chen, S., 2014. Mixotrophic cultivation of a Chlorella sorokiniana strain for enhanced biomass and lipid production. Biomass and Bioenergy 66, 204–213.

Li, Y., Ghasemi Naghdi, F., Garg, S., Adarme-Vega, T.C., Thurecht, K.J., Ghafor, W.A., Tannock, S., Schenk, P.M., 2014. A comparative study: the impact of different lipid extraction methods on current microalgal lipid research. Microb. Cell Fact. 13, 14.

Li, Y., Han, F., Xu, H., Mu, J., Chen, D., Feng, B., Zeng, H., 2014. Potential lipid accumulation and growth characteristic of the green alga Chlorella with combination cultivation mode of nitrogen (N) and phosphorus (P). Bioresour. Technol. 174, 24–32.

Li, Y., Horsman, M., Wu, N., Lan, C.Q., Dubois-Calero, N., 2008. Biofuels from microalgae. Biotechnol. Prog. 24, 815–20.

Li, Y., Lu, Z., Sun, L., Ropp, S., Kutish, G.F., Rock, D.L., Van Etten, J.L., 1997. Analysis of 74 kb of DNA located at the right end of the 330-kb chlorella virus PBCV-1 genome. Virology 237, 360–77.

Lin, H.-D., Liu, B.-H., Kuo, T.-T., Tsai, H.-C., Feng, T.-Y., Huang, C.-C., Chien, L.-F., 2013. Knockdown of PsbO leads to induction of HydA and production of photobiological H2 in the green alga Chlorella sp. DT. Bioresour. Technol. 143, 154–62.

Linz, B., Linz, a, Migunova, A., Kvitko, K., 1999. Correlation between virus-sensitivity and isoenzyme spectrum in symbiotic Chlorella-like algae. Biol. Res. 1, 76–81.

Liu, B., Benning, C., 2013. Lipid metabolism in microalgae distinguishes itself. Curr. Opin. Biotechnol. 24, 300–309.

Liu, J., Chen, F., 2016. Biology and Industrial Applications of Chlorella: Advances and

Prospects. In: Advances in Biochemical Engineering/Biotechnology. pp. 1–35.
Liu, J., Gerken, H., Li, Y., 2014a. Single-tube colony PCR for DNA amplification and transformant screening of oleaginous microalgae. J. Appl. Phycol. 26, 1719–1726.
Liu, J., Sun, Z., Gerken, H., Huang, J., Jiang, Y., Chen, F., 2014b. Genetic engineering of the green alga Chlorella zofingiensis: A modified norflurazon-resistant phytoene desaturase gene as a dominant selectable marker. Appl. Microbiol. Biotechnol. 98, 5069–5079.
Liu, L., Wang, Y., Zhang, Y., Chen, X., Zhang, P., Ma, S., 2013. Development of a New Method for Genetic Transformation of the Green Alga Chlorella ellipsoidea. Mol. Biotechnol. 54, 211–219.
Liu, W., Stewart, C.N., 2016. Plant synthetic promoters and transcription factors. Curr. Opin. Biotechnol. 37, 36–44.
Liu, W., Yuan, J.S., Stewart Jr, C.N., 2013. Advanced genetic tools for plant biotechnology. Nat. Rev. Genet. 14, 781–793.
Lohman, G.J.S., Zhang, Y., Zhelkovsky, A.M., Cantor, E.J., Evans, T.C., 2014. Efficient DNA ligation in DNA-RNA hybrid helices by Chlorella virus DNA ligase. Nucleic Acids Res. 42, 1831–1844.
Lou, S., Wang, L., He, L., Wang, Z., Wang, G., Lin, X., 2016. Production of crocetin in transgenic Chlorella vulgaris expressing genes crtRB and ZCD1. J. Appl. Phycol. 28, 1657–1665.
Lowry, J.A., Atchley, W.R., 2000. Molecular Evolution of the GATA Family of Transcription Factors: Conservation Within the DNA-Binding Domain. J. Mol. Evol. 50, 103–115.
Lu, S., Wang, J., Ma, Q., Yang, J., Li, X., Yuan, Y.-J., 2013. Phospholipid metabolism in an industry microalga Chlorella sorokiniana: the impact of inoculum sizes. PLoS One 8, e70827.
Lu, S., Wang, J., Niu, Y., Yang, J., Zhou, J., Yuan, Y., 2012. Metabolic profiling reveals growth related FAME productivity and quality of Chlorella sorokiniana with different inoculum sizes. Biotechnol. Bioeng. 109, 1651–62.
Luby-Phelps, K., 2013. The physical chemistry of cytoplasm and its influence on cell function: an update. Mol. Biol. Cell 24, 2593–6.
Lumbreras, V., Stevens, D.R., Purton, S., 1998. Efficient foreign gene expression in Chlamydomonas reinhardtii mediated by an endogenous intron. Plant J. 14, 441–447.
Lung, S.-C., Weselake, R.J., 2006. Diacylglycerol acyltransferase : A key mediator of plant triacylglycerol synthesis. Lipids 41, 1073–1088.
Luo, W., Proschold, T., Bock, C., Krienitz, L., 2010. Generic concept in Chlorella-related coccoid green algae (Chlorophyta, Trebouxiophyceae). Plant Biol. 12, 545–553.
Ma, R., Lin, X., 2014. Vitreoscilla hemoglobin gene (vgb) improves lutein production in Chlorella vulgaris. Chinese J. Oceanol. Limnol. 32, 390–396.
MacDougall, K.M., McNichol, J., McGinn, P.J., O'Leary, S.J.B., Melanson, J.E., 2011. Triacylglycerol profiling of microalgae strains for biofuel feedstock by liquid chromatography-high-resolution mass spectrometry. Anal. Bioanal. Chem. 401, 2609–2616.
Macnaghten, P., 2016. Responsible innovation and the reshaping of existing technological trajectories: the hard case of genetically modified crops. J. Responsible Innov. 3, 282–289.
Maity, J.P., Bundschuh, J., Chen, C.-Y., Bhattacharya, P., 2014. Microalgae for third generation biofuel production, mitigation of greenhouse gas emissions and wastewater treatment: Present and future perspectives – A mini review. Energy 78, 104–113.
Martins, D.A., Custódio, L., Barreira, L., Pereira, H., Ben-Hamadou, R., Varela, J., Abu-Salah, K.M., 2013. Alternative sources of n-3 long-chain polyunsaturated fatty acids in marine microalgae. Mar. Drugs 11, 2259–2281.
Maruyama, M., Horakova, I., Honda, H., Xing, X., Shiragami, N., Unno, H., 1994. Introduction of foreign DNA into Chlorella saccharophila by electroporation. Biotechnol. Tech. 8, 821–826.

Mata, T.M., Martins, A.A., Caetano, N.S., 2010. Microalgae for biodiesel production and other applications: A review. Renew. Sustain. Energy Rev. 14, 217–232.

Matt, G.Y., Umen, J.G., 2018. Cell-Type Transcriptomes of the Multicellular Green Alga Volvox carteri Yield Insights into the Evolutionary Origins of Germ and Somatic Differentiation Programs. G3 Genes, Genomes, Genet. 8, 531–550.

McCullen, C.A., Binns, A.N., 2006. *Agrobacterium tumefaciens* and Plant Cell Interactions and Activities Required for Interkingdom Macromolecular Transfer. Annu. Rev. Cell Dev. Biol. 22, 101–127.

McKie-Krisberg, Z.M., Laurens, L.M.L., Huang, A., Polle, J.E.W., 2018. Comparative energetics of carbon storage molecules in green algae. Algal Res. 31, 326–333.

McKinney, M.K., Cravatt, B.F., 2005. Structure and Function of Fatty Acid Amide Hydrolase. Annu. Rev. Biochem. 74, 411–432.

Meints, R.H., Van Etten, J.L., Kuczmarski, D., Lee, K., Ang, B., 1981. Viral infection of the symbiotic chlorella-like alga present in Hydra viridis. Virology 113, 698–703.

Merchant, S.S., Kropat, J., Liu, B., Shaw, J., Warakanont, J., 2012. TAG, You're it! Chlamydomonas as a reference organism for understanding algal triacylglycerol accumulation. Curr. Opin. Biotechnol. 23, 352–363.

Merika, M., Orkin, S.H., 1993. DNA-binding specificity of GATA family transcription factors. Mol. Cell. Biol. 13, 3999–4010.

Meyers, B., Zaltsman, A., Lacroix, B., Kozlovsky, S. V., Krichevsky, A., 2010. Nuclear and plastid genetic engineering of plants: comparison of opportunities and challenges. Biotechnol. Adv. 28, 747–56.

Miao, X., Wu, Q., 2006. Biodiesel production from heterotrophic microalgal oil. Bioresour. Technol. 97, 841–846.

Miller, P.W., Russell, B.L., Schmidt, R.R., 1994. Transcription Initiation Site of a NADP-Specific Glutamate Dehydrogenase Gene and Potential Use of Its Promoter Region to Express Foreign Genes in Ammonium-Cultured Chlorella sorokiniana Cells. J. Appl. Phycol. 6, 211–223.

Minhas, A.K., Hodgson, P., Barrow, C.J., Adholeya, A., 2016. A review on the assessment of stress conditions for simultaneous production of microalgal lipids and carotenoids. Front. Microbiol. 7, 1–19.

Mini, P., Demurtas, O.C., Valentini, S., Pallara, P., Aprea, G., Ferrante, P., Giuliano, G., 2018. Agrobacterium-mediated and electroporation-mediated transformation of Chlamydomonas reinhardtii: a comparative study. BMC Biotechnol. 18, 11.

Mitra, A., Higgins, D.W., 1994. The Chlorella virus adenine methyltransferase gene promoter is a strong promoter in plants. Plant Mol. Biol. 26, 85–93.

Mitra, A., Higgins, D.W., Rohe, N.J., 1994. A Chlorella Virus Gene Promoter Functions as a Strong Promoter Both in Plants and Bacteria. Biochem. Biophys. Res. Commun. 204, 187–194.

Moreno-Garcia, L., Adjallï¿½, K., Barnabï¿½, S., Raghavan, G.S.V., 2017. Microalgae biomass production for a biorefinery system: Recent advances and the way towards sustainability. Renew. Sustain. Energy Rev. 76, 493–506.

Morschett, H., Wiechert, W., Oldiges, M., 2016. Automation of a Nile red staining assay enables high throughput quantification of microalgal lipid production. Microb. Cell Fact. 15, 1–11.

Mühlroth, A., Li, K., Røkke, G., Winge, P., Olsen, Y., Hohmann-Marriott, M., Vadstein, O., Bones, A., 2013. Pathways of Lipid Metabolism in Marine Algae, Co-Expression Network, Bottlenecks and Candidate Genes for Enhanced Production of EPA and DHA in Species of Chromista. Mar. Drugs 11, 4662–4697.

Mujtaba, S., Manzur, K.L., Gurnon, J.R., Kang, M., Van Etten, J.L., Zhou, M.-M., 2008. Epigenetic transcriptional repression of cellular genes by a viral SET protein. Nat. Cell Biol. 10, 1114–22.

Müller, F.W., Igloi, G.L., Beck, C.F., 1992. Structure of a gene encoding heat-shock protein HSP70 from the unicellular alga Chlamydomonas reinhardtii. Gene 111, 165–173.

Mushegian, A.R., Shepherd, R.J., 1995. Genetic elements of plant viruses as tools for genetic engineering. Microbiol. Rev. 59, 548–578.

NAABB, 2014. National Alliance for Advanced Biofuels and Bioproducts (NAABB) final report [WWW Document]. URL http://www.energy.gov/eere/bioenergy/downloads/national-alliance-advanced-biofuels-and-bioproducts-synopsis-naabb-final (accessed 3.12.16).

Naik, S.N., Goud, V. V., Rout, P.K., Dalai, A.K., 2010. Production of first and second generation biofuels: A comprehensive review. Renew. Sustain. Energy Rev. 14, 578–597.

Nakagawa, S., Niimura, Y., Gojobori, T., Tanaka, H., Miura, K. ichiro, 2008. Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes. Nucleic Acids Res. 36, 861–871.

Nakamura, Y., Gojobori, T., Ikemura, T., 1999. Codon usage tabulated from the international DNA sequence databases; its status 1999. Nucleic Acids Res. 27, 292.

NCABB, 2016. Nebraska Coalition for Algal Biology and Biotechnology: Databases [WWW Document]. Nebraska Coalit. Algal Biol. Biotechnol. URL http://www.unl.edu/ncabb/databases (accessed 3.12.16).

Nebraska Center for Virology, T., Van Etten Lab, T., 2012. Isolate Chlorella Virus [WWW Document]. World of Chlorella Viruses. URL http://ncv.unl.edu/vanettenlab/isolate.html (accessed 2.29.16).

Negi, S., Barry, A.N., Friedland, N., Sudasinghe, N., Subramanian, S., Pieris, S., Holguin, F.O., Dungan, B., Schaub, T., Sayre, R., 2015. Impact of nitrogen limitation on biomass, photosynthesis, and lipid accumulation in Chlorella sorokiniana. J. Appl. Phycol.

Neofotis, P., Huang, A., Sury, K., Chang, W., Joseph, F., Gabr, A., Twary, S., Qiu, W., Holguin, O., Polle, J.E.W., 2016. Characterization and classification of highly productive microalgae strains discovered for biofuel and bioproduct generation. Algal Res. 15, 164–178.

Ng, S.L., Harikrishna, J.A., Abu Bakar, F., Yeo, C.C., Cha, T.S., 2016. Heterologous expression of the Streptococcus pneumoniae yoeB and pezT toxin genes is lethal in Chlorella vulgaris. Algal Res. 19, 21–29.

Nguyen, P.-S., Falcone, D.L., Graves, M. V., 2009. The A312L 5'-UTR of Chlorella virus PBCV-1 is a translational enhancer in Arabidopsis thaliana. Virus Res. 140, 138–46.

Nichols, H.W., Bold, H.C., 1965. Trichosarcina polymorpha Gen. et Sp. Nov. J. Phycol. 1, 34–38.

Nickolai, D.J., Lammel, C.J., Byford, B.A., Morris, J.H., Kaplan, E.B., Hadley, W.K., Brooks, G.F., 1985. Effects of storage temperature and pH on the stability of eleven beta-lactam antibiotics in MIC trays. J. Clin. Microbiol. 21, 366–370.

Niu, Y.F., Zhang, M.H., Xie, W.H., Li, J.N., Gao, Y.F., Yang, W.D., Liu, J.S., Li, H.Y., 2011. A new inducible expression system in a transformed green alga, Chlorella vulgaris. Genet. Mol. Res. 10, 3427–3434.

Nymark, M., Sharma, A.K., Sparstad, T., Bones, A.M., Winge, P., 2016. A CRISPR/Cas9 system adapted for gene editing in marine algae. Sci. Rep. 6, 24951.

Odell, M., Malinina, L., Sriskanda, V., Teplova, M., Shuman, S., 2003. Analysis of the DNA joining repertoire of Chlorella virus DNA ligase and a new crystal structure of the ligase-adenylate intermediate. Nucleic Acids Res. 31, 5090–5100.

Olmstead, I.L.D., Hill, D.R.A., Dias, D.A., Jayasinghe, N.S., Callahan, D.L., Kentish, S.E., Scales, P.J., Martin, G.J.O., 2013. A quantitative analysis of microalgal lipids for optimization of biodiesel and omega-3 production. Biotechnol. Bioeng. 110, 2096–2104.

Ördög, V., Stirk, W.A., Bálint, P., Aremu, A.O., Okem, A., Lovász, C., Molnár, Z., van Staden, J., 2016. Effect of temperature and nitrogen concentration on lipid productivity and fatty acid composition in three Chlorella strains. Algal Res. 16, 141–149.

Ortiz-Matamoros, M.F., Islas-Flores, T., Voigt, B., Menzel, D., Baluška, F., Villanueva, M.A.,

2015. Heterologous DNA Uptake in Cultured Symbiodinium spp. Aided by Agrobacterium tumefaciens. PLoS One 10, 1–16.

Otsuka, H., 1961. CHANGES OF LIPID AND CARBOHYDRATE CONTENTS IN CHLORELLA CELLS DURING THE SULFUR STARVATION, AS STUDIED BY THE TECHNIQUE OF SYNCHRONOUS CULTURE. J. Gen. Appl. Microbiol. 7, 72–77.

Packeiser, H., Lim, C., Balagurunathan, B., Wu, J., Zhao, H., 2013. An extremely simple and effective colony PCR procedure for bacteria, yeasts, and microalgae. Appl. Biochem. Biotechnol. 169, 695–700.

Pacurar, D.I., Thordal-Christensen, H., Pacurar, M.L., Pamfil, D., Botez, M.L., Bellini, C., 2011. Agrobacterium tumefaciens: From crown gall tumors to genetic transformation. Physiol. Mol. Plant Pathol. 76, 76–81.

Padmanabhan, S., Banerjee, S., Mandi, N., 2011. Screening of Bacterial Recombinants: Strategies and Preventing False Positives. In: Brown, G.G. (Ed.), Molecular Cloning - Selected Applications in Medicine and Biology. Rijeka.

Paranjape, K., Leite, G.B., Hallenbeck, P.C., 2016. Effect of nitrogen regime on microalgal lipid production during mixotrophic growth with glycerol. Bioresour. Technol. 214, 778–786.

Park, H.-H., Choi, T.-J., 2004. Application of a Promoter Isolated from Chlorella Virus in Chlorella Transformation System. Plant Pathol. J. 20, 158–163.

Park, H.-J., Yoon, H.-M., Jung, H.-K., Choi, T.-J., 2005. Isolation and Characterization o Chlorella Virus from Fresh Water in Korea and Application in Chlorella Transformation System. Plant Pathol. J. 21, 13–20.

Patron, N.J., Orzaez, D., Marillonnet, S., Warzecha, H., Matthewman, C., Youles, M., Raitskin, O., Leveau, A., Farre, G., Rogers, C., Smith, A., Hibberd, J., Webb, A.A.R., Locke, J., Schornack, S., Ajioka, J., Baulcombe, D.C., Zipfel, C., Kamoun, S., Jones, J.D.G., Kuhn, H., Robatzek, S., Esse, H.P. Van, Sanders, D., Oldroyd, G., Martin, C., Field, R., O'Connor, S., Fox, S., Wulff, B., Miller, B., Breakspear, A., Radhakrishnan, G., Delaux, P.-M., Loque, D., Granell, A., Tissier, A., Shih, P., Brutnell, T.P., Quick, W.P., Rischer, H., Fraser, P.D., Aharoni, A., Raines, C., South, P.F., Ane, J.-M., Hamberger, B.R., Langdale, J., Stougaard, J., Bouwmeester, H., Udvardi, M., Murray, J.A.H., Ntoukakis, V., Schafer, P., Denby, K., Edwards, K.J., Osbourn, A., Haseloff, J., 2015. Standards for Plant Synthetic Biology: A Common Syntax for Exchange of DNA Parts. New Phytol. 13–19.

Pauli, J.N., Mendoza, J.E., Steffan, S.A., Carey, C.C., Weimer, P.J., Peery, M.Z., 2014. A syndrome of mutualism reinforces the lifestyle of a sloth. Proc. R. Soc. B Biol. Sci. 281.

Pereira, H., Barreira, L., Figueiredo, F., Custódio, L., Vizetto-Duarte, C., Polo, C., Rešek, E., Engelen, A., Varela, J., 2012. Polyunsaturated Fatty acids of marine macroalgae: potential for nutritional and pharmaceutical applications. Mar. Drugs 10, 1920–35.

Perez-Garcia, O., Escalante, F.M.E., de-Bashan, L.E., Bashan, Y., 2011. Heterotrophic cultures of microalgae: Metabolism and potential products. Water Res. 45, 11–36.

Petkov, G., Garcia, G., 2007. Which are fatty acids of the green alga Chlorella? Biochem. Syst. Ecol. 35, 281–285.

Petro, T.M., Agarkova, I. V, Zhou, Y., Yolken, R.H., Van Etten, J.L., Dunigan, D.D., 2015. Response of Mammalian Macrophages to Challenge with the Chlorovirus Acanthocystis turfacea Chlorella Virus 1. J Virol 89, 12096–12107.

Pick, U., Rachutin-Zalogin, T., 2012. Kinetic anomalies in the interactions of Nile red with microalgae. J. Microbiol. Methods 88, 189–196.

Pires, J.C.M., 2017. COP21: The algae opportunity? Renew. Sustain. Energy Rev. 79, 867–877.

Prasad, B., Vadakedath, N., Jeong, H.-J., General, T., Cho, M.-G., Lein, W., 2014. Agrobacterium tumefaciens-mediated genetic transformation of haptophytes (Isochrysis species). Appl. Microbiol. Biotechnol. 98, 8629–8639.

Pratheesh, P.T., Shonima, G.M., Thomas, J., Abraham, C.I., Muraleedhara, K.G., 2012. Study on efficacy of different Agrobacterium tumefaciens strains in genetic transformation of microalga Chlamydomonas reinhardtii. Pelagia Res. Libr. 3, 2679–2686.

Pratheesh, P.T., Vineetha, M., Kurup, G.M., 2014. An efficient protocol for the Agrobacterium-mediated genetic transformation of microalga Chlamydomonas reinhardtii. Mol. Biotechnol. 56, 507–515.

Přibyl, P., Cepák, V., Zachleder, V., 2012. Production of lipids in 10 strains of chlorella and parachlorella, and enhanced lipid productivity in chlorella vulgaris. Appl. Microbiol. Biotechnol. 94, 549–561.

Proschold, T., Darienko, T., Silva, P.C., Reisser, W., Krienitz, L., 2011. The systematics of Zoochlorella revisited employing an integrative approach. Environ. Microbiol. 13, 350–364.

Puchta, H., 2003. Marker-free transgenic plants. Plant Cell. Tissue Organ Cult. 74, 123–134.

Qi, C., Sun, J., Xia, Y., Yu, R., Wei, W., Xiang, J., Jin, Q., Xiao, H., Wang, X., 2018. Fatty Acid Profile and the sn-2 Position Distribution in Triacylglycerols of Breast Milk during Different Lactation Stages. J. Agric. Food Chem. 66, 3118–3126.

Qin, S., Lin, H., Jiang, P., 2012. Advances in genetic engineering of marine algae. Biotechnol. Adv. 30, 1602–1613.

Quax, T.E.F., Claassens, N.J., Söll, D., van der Oost, J., 2015. Codon Bias as a Means to Fine-Tune Gene Expression. Mol. Cell 59, 149–161.

Radakovits, R., Jinkerson, R.E., Darzins, A., Posewitz, M.C., 2010. Genetic engineering of algae for enhanced biofuel production. Eukaryot. Cell 9, 486–501.

Radzun, K.A., Wolf, J., Jakob, G., Zhang, E., Stephens, E., Ross, I., Hankamer, B., 2015. Automated nutrient screening system enables high-throughput optimisation of microalgae production conditions. Biotechnol. Biofuels 8, 1–17.

Raha, H.E., Shafii, M.B., Roshandel, R., 2018. Energy efficient cultivation of microalgae using phosphorescence materials and mirrors. Sustain. Cities Soc. 41, 449–454.

Rai, M.P., Gautom, T., Sharma, N., 2015. Effect of Salinity, pH, Light Intensity on Growth and Lipid Production of Microalgae for Bioenergy Application. Online J. Biol. Sci. 15, 260–267.

Rajam, M. V, Kumar, S.V., 2007. Green Alga (Chlamydomonas reinhardtii). In: Wang, K. (Ed.), Methods in Molecular Biology: Agrobacterium Protocols Volume 2. Springer, pp. 421–33.

Rakesh, S., Dhar, D.W., Prasanna, R., Saxena, A.K., Saha, S., Shukla, M., Sharma, K., 2015. Cell disruption methods for improving lipid extraction efficiency in unicellular microalgae. Eng. Life Sci. 15, 443–447.

Rakkhumkaew, N., Kawasaki, T., Fujie, M., Yamada, T., 2013. Prolonged synthesis of hyaluronan by Chlorella cells infected with chloroviruses. J. Biosci. Bioeng. 115, 527–531.

Ramos, M.J., Fernández, C.M., Casas, A., Rodríguez, L., Pérez, Á., 2009. Influence of fatty acid composition of raw materials on biodiesel properties. Bioresour. Technol. 100, 261–268.

Rasala, B. a, Barrera, D.J., Ng, J., Plucinak, T.M., Rosenberg, J.N., Weeks, D.P., Oyler, G.A., Peterson, T.C., Haerizadeh, F., Mayfield, S.P., 2013. Expanding the spectral palette of fluorescent proteins for the green microalga Chlamydomonas reinhardtii. Plant J. 74, 545–56.

Rasala, B.A., Mayfield, S.P., 2015. Photosynthetic biomanufacturing in green algae; production of recombinant proteins for industrial, nutritional, and medical uses. Photosynth. Res. 123, 227–239.

Rathod, J.P., Prakash, G., Pandit, R., Lali, A.M., 2013. Agrobacterium-mediated transformation of promising oil-bearing marine algae Parachlorella kessleri. Photosynth. Res. 118, 141–146.

Rawat, I., Ranjith Kumar, R., Mutanda, T., Bux, F., 2013. Biodiesel from microalgae: A critical evaluation from laboratory to large scale production. Appl. Energy 103, 444–467.

Reddy, P.H., Johnson, A.M.A., Kumar, J.K., Naveen, T., Devi, M.C., 2017. Heterologous

expression of Infectious bursal disease virus VP2 gene in Chlorella pyrenoidosa as a model system for molecular farming. Plant Cell, Tissue Organ Cult. 131, 119–126.

Reisser, W., Burbank, D.E., Meints, S.M., Meints, R.H., Becker, B., Van Etten, J.L., 1988. A comparison of viruses infecting two different Chlorella-like green Algae. Virology 167, 143–149.

Ren, X., Chen, J., Deschênes, J.S., Tremblay, R., Jolicoeur, M., 2016. Glucose feeding recalibrates carbon flux distribution and favours lipid accumulation in Chlorella protothecoides through cell energetic management. Algal Res. 14, 83–91.

Roberto Mantovani, 1998. A survey of 178 NF-Y binding CCAAT boxes. Nucleic Acids Res. 26, 1135–1143.

Robertson, R., Guihéneuf, F., Bahar, B., Schmid, M., Stengel, D., Fitzgerald, G., Ross, R., Stanton, C., 2015. The Anti-Inflammatory Effect of Algae-Derived Lipid Extracts on Lipopolysaccharide (LPS)-Stimulated Human THP-1 Macrophages. Mar. Drugs 13, 5402–5424.

Rolloos, M., Hooykaas, P.J.J., van der Zaal, B.J., 2015. Enhanced targeted integration mediated by translocated I-SceI during the Agrobacterium mediated transformation of yeast. Sci. Rep. 5, 8345.

Romani, G., Piotrowski, A., Hillmer, S., Gurnon, J., Van Etten, J.L., Moroni, A., Thiel, G., Hertel, B., 2013. A virus-encoded potassium ion channel is a structural protein in the chlorovirus Paramecium bursaria chlorella virus 1 virion. J. Gen. Virol. 94, 2549–56.

Rommens, C.M., 2004. All-native DNA transformation: A new approach to plant genetic engineering. Trends Plant Sci. 9, 457–464.

Rosenberg, J.N., Kobayashi, N., Barnes, A., Noel, E. a, Betenbaugh, M.J., Oyler, G. a, 2014. Comparative analyses of three Chlorella species in response to light and sugar reveal distinctive lipid accumulation patterns in the Microalga C. sorokiniana. PLoS One 9, e92460.

Rowe, J.M., Jeanniard, A., Gurnon, J.R., Xia, Y., Dunigan, D.D., Van Etten, J.L., Blanc, G., 2014. Global analysis of Chlorella variabilis NC64A mRNA profiles during the early phase of Paramecium bursaria chlorella virus-1 infection. PLoS One 9, e90988.

Roy, A.L., Singer, D.S., 2015. Core promoters in transcription: Old problem, new insights. Trends Biochem. Sci. 40, 165–171.

Ruecker, O., Zillner, K., Groebner-Ferreira, R., Heitzer, M., 2008. Gaussia-luciferase as a sensitive reporter gene for monitoring promoter activity in the nucleus of the green alga Chlamydomonas reinhardtii. Mol. Genet. Genomics 280, 153–162.

Ruiz-Lopez, N., Haslam, R.P., Napier, J.A., Sayanova, O., 2014a. Successful high-level accumulation of fish oil omega-3 long-chain polyunsaturated fatty acids in a transgenic oilseed crop. Plant J. 77, 198–208.

Ruiz-Lopez, N., Usher, S., Sayanova, O. V., Napier, J.A., Haslam, R.P., 2014b. Modifying the lipid content and composition of plant seeds: engineering the production of LC-PUFA. Appl. Microbiol. Biotechnol. 99, 143–154.

Rumin, J., Bonnefond, H., Saint-Jean, B., Rouxel, C., Sciandra, A., Bernard, O., Cadoret, J.P., Bougaran, G., 2015. The use of fluorescent Nile red and BODIPY for lipid measurement in microalgae. Biotechnol. Biofuels 8, 1–16.

Ryckebosch, E., Bruneel, C., Muylaert, K., Foubert, I., 2012. Microalgae as an alternative source of omega-3 long chain polyunsaturated fatty acids. Lipid Technol. 24, 128–130.

Ryckebosch, E., Bruneel, C., Termote-Verhalle, R., Goiris, K., Muylaert, K., Foubert, I., 2014. Nutritional evaluation of microalgae oils rich in omega-3 long chain polyunsaturated fatty acids as an alternative for fish oil. Food Chem. 160, 393–400.

Safi, C., Zebib, B., Merah, O., Pontalier, P.Y., Vaca-Garcia, C., 2014. Morphology, composition, production, processing and applications of Chlorella vulgaris: A review. Renew. Sustain. Energy Rev. 35, 265–278.

Sajjadi, B., Chen, W.-Y., Raman, A.A.A., Ibrahim, S., 2018. Microalgae lipid and biomass for

biofuel production: A comprehensive review on lipid enhancement strategies and their effects on fatty acid composition. Renew. Sustain. Energy Rev. 97, 200–232.

SAMS, 2018. Culture Collection of Algae and Protozoa [WWW Document]. URL https://www.ccap.ac.uk/advanced-search.php (accessed 9.5.18).

Sanitha, M., Radha, S., Fatima, A.A., Devi, S.G., Ramya, M., 2014. Agrobacterium-mediated transformation of three freshwater microalgal strains. Polish J. Microbiol. 63, 387–392.

Santini, A., Cammarata, S.M., Capone, G., Ianaro, A., Tenore, G.C., Pani, L., Novellino, E., 2018. Nutraceuticals: opening the debate for a regulatory framework. Br. J. Clin. Pharmacol. 84, 659–672.

Sarrafzadeh, M.H., La, H.-J., Lee, J.-Y., Cho, D.-H., Shin, S.-Y., Kim, W.-J., Oh, H.-M., 2014. Microalgae biomass quantification by digital image processing and RGB color analysis. J. Appl. Phycol. 27, 205–209.

Sarrion-Perdigones, A., Falconi, E.E., Zandalinas, S.I., Juárez, P., Fernández-del-Carmen, A., Granell, A., Orzaez, D., 2011. GoldenBraid: An iterative cloning system for standardized assembly of reusable genetic modules. PLoS One 6.

Sarrion-Perdigones, A., Vazquez-Vilar, M., Palaci, J., Castelijns, B., Forment, J., Ziarsolo, P., Blanca, J., Granell, A., Orzaez, D., 2013. GoldenBraid 2.0: A Comprehensive DNA Assembly Framework for Plant Synthetic Biology. Plant Physiol. 162, 1618–1631.

Sathasivam, R., Radhakrishnan, R., Hashem, A., Abd_Allah, E.F., 2017. Microalgae metabolites: A rich source for food and medicine. Saudi J. Biol. Sci.

Saunders, K., Lomonossoff, G.P., 2013. Exploiting plant virus-derived components to achieve in planta expression and for templates for synthetic biology applications. New Phytol. 200, 16–26.

Sayanova, O., Mimouni, V., Ulmann, L., Morant-Manceau, A., Pasquet, V., Schoefs, B., Napier, J.A., 2017. Modulation of lipid biosynthesis by stress in diatoms. Philos. Trans. R. Soc. B Biol. Sci. 372.

Sayre, R., 2010. Microalgae: The Potential for Carbon Capture. Bioscience 60, 722–727.

Scaife, M.A., Smith, A.G., 2016. Towards developing algal synthetic biology. Biochem. Soc. Trans. 44, 716–722.

Scala, S., Bowler, C., 2001. Molecular insights into the novel aspects of diatom biology. Cell. Mol. Life Sci. 58, 1666–1673.

Scherholz, M.L., Curtis, W.R., 2013. Achieving pH control in microalgal cultures through fed-batch addition of stoichiometrically-balanced growth media. BMC Biotechnol. 13, 39.

Schroda, M., Beck, C.F., Vallon, O., 2002. Sequence elements within an HSP70 promoter counteract transcriptional transgene silencing in Chlamydomonas. Plant J. 31, 445–455.

Schroda, M., Blöcker, D., Beck, C.F., 2000. The HSP70A promoter as a tool for the improved expression of transgenes in Chlamydomonas. Plant J. 21, 121–131.

Schüler, L.M., Schulze, P.S.C., Pereira, H., Barreira, L., León, R., Varela, J., 2017. Trends and strategies to enhance triacylglycerols and high-value compounds in microalgae. Algal Res. 25, 263–273.

Schuster, A.M., Graves, M., Korth, K., Ziegelbein, M., Brumbaugh, J., Grone, D., Meints, R.H., 1990. Transcription and sequence studies of a 4.3-kbp fragment from a ds-DNA eukaryotic algal virus. Virology 176, 515–523.

Scranton, M.A., Ostrand, J.T., Georgianna, D.R., Lofgren, S.M., Li, D., Ellis, R.C., Carruthers, D.N., Dräger, A., Masica, D.L., Mayfield, S.P., 2016. Synthetic promoters capable of driving robust nuclear gene expression in the green alga Chlamydomonas reinhardtii. Algal Res. 15, 135–142.

Selvaraj, V., Singh, H., Ramaswamy, S., 2013. Chlorella-Induced Psychosis. Psychosomatics 54, 303–304.

Shahmuradov, I.A., Gammerman, A.J., Hancock, J.M., Bramley, P.M., Solovyev, V. V., 2003. PlantProm: A database of plant promoter sequences. Nucleic Acids Res. 31, 114–117.

Sharif, N., Munir, N., Saleem, F., Naz, S., 2015. Factors Affecting Agrobacterium Mediated

Transformation Of Indigenous Chlorella Vulgaris Bayerinck. Bangladesh J. Bot. 44, 323–326.

Shihira, I., Krauss, R.W., 1965. Chlorella: Physiology and Taxonomy of Forty-one Isolates. University of Maryland.

Shin, S.E., Lim, J.M., Koh, H.G., Kim, E.K., Kang, N.K., Jeon, S., Kwon, S., Shin, W.S., Lee, B., Hwangbo, K., Kim, J., Ye, S.H., Yun, J.Y., Seo, H., Oh, H.M., Kim, K.J., Kim, J.S., Jeong, W.J., Chang, Y.K., Jeong, B.R., 2016. CRISPR/Cas9-induced knockout and knock-in mutations in Chlamydomonas reinhardtii. Sci. Rep. 6, 1–15.

Show, P.L., Tang, M.S.Y., Nagarajan, D., Ling, T.C., Ooi, C.W., Chang, J.S., 2017. A holistic approach to managing microalgae for biofuel applications. Int. J. Mol. Sci. 18.

Shrestha, R.P., Haerizadeh, F., Hildebrand, M., 2013. Molecular Genetic Manipulation of Microalgae : Principles and Applications. In: Richmond, A., Hu, Q. (Eds.), Handbook of Microalgal Culture: Applied Phycology and Biotechnology. Blackwell Publishing Ltd., pp. 146–167.

Sili, C., Torzillo, G., Vonshak, A., 2012. Arthrospira (Spirulina). In: Whitton, B.A. (Ed.), Ecology of Cyanobacteria II: Their Diversity in Space and Time. Springer Science+Business Media B.V., pp. 677–705.

Simon, D.P., Anila, N., Gayathri, K., Sarada, R., 2016. Heterologous expression of β-carotene hydroxylase in Dunaliella salina by Agrobacterium-mediated genetic transformation. Algal Res. 18, 257–265.

Slocombe, S.P., Zhang, Q., Black, K.D., Day, J.G., Stanley, M.S., 2013. Comparison of screening methods for high-throughput determination of oil yields in micro-algal biofuel strains. J. Appl. Phycol. 25, 961–972.

Slocombe, S.P., Zhang, Q., Ross, M., Anderson, A., Thomas, N.J., Lapresa, Á., Rad-Menéndez, C., Campbell, C.N., Black, K.D., Stanley, M.S., Day, J.G., 2015. Unlocking nature's treasure-chest: screening for oleaginous algae. Sci. Rep. 5, 9844.

Smale, S.T., Baltimore, D., 1989. The 'initiator' as a transcription control element. Cell 57, 103–113.

Sorensen, I., Fei, Z., Andreas, A., Willats, W.G.T., Domozych, D.S., Rose, J.K.C., 2014. Stable transformation and reverse genetic analysis of Penium margaritaceum: A platform for studies of charophyte green algae, the immediate ancestors of land plants. Plant J. 77, 339–351.

Soria-Guerra, R.E., Ramírez-Alonso, J.I., Ibáñez-Salazar, A., Govea-Alonso, D.O., Paz-Maldonado, L.M.T., Bañuelos-Hernández, B., Korban, S.S., Rosales-Mendoza, S., 2014. Expression of an HBcAg-based antigen carrying angiotensin II in Chlamydomonas reinhardtii as a candidate hypertension vaccine. Plant Cell. Tissue Organ Cult. 116, 133–139.

Sorokin, C., 1959. Tabular Comparative Data for the Low- and High- Temperature Strains of Chlorella. Nature 184, 613–614.

Sorokin, C., Krauss, R.W., 1959. Maximum growth rates of Chlorella in steady-state and in synchronized cultures. Proc. Natl. Acad. Sci. U. S. A. 45, 1740–4.

Sorokin, C., Myers, J., 1953. A High-Temperature Strain of Chlorella. Science (80-.). 117, 330–331.

Specht, E., Miyake-Stoner, S., Mayfield, S., 2010. Micro-algae come of age as a platform for recombinant protein production. Biotechnol. Lett. 32, 1373–83.

Spicer, A., Purton, S., 2016. Genetic Engineering of Microalgae: Current Status and Future Prospects. In: Slocombe, S.P., Benemann, J.R. (Eds.), Microalgal Production for Biomass and High-Value Products. Taylor & Francis Group, pp. 139–164.

Srinivasan, R., Gothandam, K.M., 2016. Synergistic action of D-glucose and acetosyringone on agrobacterium strains for efficient dunaliella transformation. PLoS One 11, 1–13.

Stanley, M., Jenkins, T., Schlarb-Ridley, B., Parker, B., 2013. A UK Roadmap for Algal Technologies, Algal Bioenergy Special Interest Group.

Stevens, D.R., Purton, S., 1997. Genetic Engineering of Eukarygtic Algae: Progress and Prospects. J. Phycol. 33, 713–722.

Stevens, D.R., Rochaix, J.D., Purton, S., 1996. The bacterial phleomycin resistance gene ble as a dominant selectable marker in Chlamydomonas. Mol. Gen. Genet. 251, 23–30.

Su, Y., Song, K., Zhang, P., Su, Y., Cheng, J., Chen, X., 2017. Progress of microalgae biofuel's commercialization. Renew. Sustain. Energy Rev. 74, 402–411.

Sueoka, N., 1960. Mitotic Replication of Deoxyribonucleic Acid in Chlamydomonas Reinhardi. Proc. Natl. Acad. Sci. U. S. A. 46, 83–91.

Takeda, H., 1988. Classification of Chlorella strains by cell wall sugar composition. Phytochemistry 27, 3823–3826.

Tan, W., Adebusuyi, A.A., 2016. Proton and nutrient balanced medium for scalable, practical pH control in high-density Chlorella vulgaris cultures. Algal Res. 16, 119–126.

Tang, Y., Rosenberg, J.N., Bohutskyi, P., Yu, G., Betenbaugh, M.J., Wang, F., 2016a. Microalgae as a Feedstock for Biofuel Production: Green Fuels and Golden Opportunities. BioResources 11.

Tang, Y., Zhang, Y., Rosenberg, J.N., Sharif, N., Betenbaugh, M.J., Wang, F., 2016b. Efficient lipid extraction and quantification of fatty acids from algal biomass using accelerated solvent extraction (ASE). RSC Adv. 6, 29127–29134.

Thanh, T., Chi, V.T.Q., Omar, H., Abdullah, M.P., Napis, S., 2012. Sequence analysis and potentials of the native RbcS promoter in the development of an alternative eukaryotic expression system using green microalga ankistrodesmus convolutus. Int. J. Mol. Sci. 13, 2676–2691.

Thiriet-Rupert, S., Carrier, G., Chénais, B., Trottier, C., Bougaran, G., Cadoret, J.P., Schoefs, B., Saint-Jean, B., 2016. Transcription factors in microalgae: Genome-wide prediction and comparative analysis. BMC Genomics 17, 1–16.

Tirichine, L., Bowler, C., 2011. Decoding algal genomes: tracing back the history of photosynthetic life on Earth. Plant J. 66, 45–57.

Tischner, R., Lorenzen, H., 1979. Nitrate uptake and nitrate reduction in synchronous Chlorella. Planta 146, 287–292.

Tzfira, T., Citovsky, V., 2006. Agrobacterium-mediated genetic transformation of plants: biology and biotechnology. Curr. Opin. Biotechnol. 17, 147–154.

Tzfira, T., Frankman, L.R., Vaidya, M., Citovsky, V., 2003. Site-Specific Integration of Agrobacterium tumefaciens T-DNA via Double-Stranded Intermediates. Plant Physiol. 133, 1011–1023.

Ubeda-Minguez, P., Chileh, T., Dautor, Y., Garcia-Maroto, F., Alonso, D.L., 2015. Tools for microalgal biotechnology: development of an optimized transformation method for an industrially promising microalga—Tetraselmis chuii. J. Appl. Phycol. 27, 223–232.

Úbeda-Mínguez, P., García-Maroto, F., Alonso, D.L., 2017. Heterologous expression of DGAT genes in the marine microalga Tetraselmis chui leads to an increase in TAG content. J. Appl. Phycol. 29, 1913–1926.

Ulker, B., Li, Y., Rosso, M.G., Logemann, E., Somssich, I.E., Weisshaar, B., 2008. T-DNA-mediated transfer of Agrobacterium tumefaciens chromosomal DNA into plants. Nat. Biotechnol. 26, 1015–1017.

Unkefer, C.J., Sayre, R.T., Magnuson, J.K., Anderson, D.B., Baxter, I., Blaby, I.K., Brown, J.K., Carleton, M., Cattolico, R.A., Dale, T., Devarenne, T.P., Downes, C.M., Dutcher, S.K., Fox, D.T., Goodenough, U., Jaworski, J., Holladay, J.E., Kramer, D.M., Koppisch, A.T., Lipton, M.S., Marrone, B.L., McCormick, M., Molnár, I., Mott, J.B., Ogden, K.L., Panisko, E.A., Pellegrini, M., Polle, J., Richardson, J.W., Sabarsky, M., Starkenburg, S.R., Stormo, G.D., Teshima, M., Twary, S.N., Unkefer, P.J., Yuan, J.S., Olivares, J.A., 2017. Review of the algal biology program within the National Alliance for Advanced Biofuels and Bioproducts. Algal Res. 22, 187–215.

Vaezi, R., Napier, J.A., Sayanova, O., 2013. Identification and functional characterization of

genes encoding omega-3 polyunsaturated fatty acid biosynthetic activities from unicellular microalgae. Mar. Drugs 11, 5116–29.

Van Etten, J.L., Agarkova, I., Dunigan, D.D., Tonetti, M., De Castro, C., Duncan, G.A., 2017. Chloroviruses have a sweet tooth. Viruses 9, 1–23.

Van Etten, J.L., Burbank, D.E., Schuster, A.M., Meints, R.H., 1985a. Lytic viruses infecting a chlorella-like alga. Virology 140, 135–143.

Van Etten, J.L., Dunigan, D.D., 2012. Chloroviruses: not your everyday plant virus. Trends Plant Sci. 17, 1–8.

Van Etten, J.L., Gurnon, J.R., Yanai-Balser, G.M., Dunigan, D.D., Graves, M. V, 2010. Chlorella viruses encode most, if not all, of the machinery to glycosylate their glycoproteins independent of the endoplasmic reticulum and Golgi. Biochim. Biophys. Acta 1800, 152–9.

Van Etten, J.L., Meints, R.H., Kuczmarski, D., Burbank, D.E., Lee, K., 1982. Viruses of symbiotic Chlorella-like algae isolated from Paramecium bursaria and Hydra viridis. Proc. Natl. Acad. Sci. U. S. A. 79, 3867–71.

Van Etten, J.L., Van Etten, C.H., Johnson, J.K., Burbank, D.E., 1985b. A survey for viruses from fresh water that infect a eucaryotic chlorella-like green alga. Appl. Environ. Microbiol. 49, 1326–1328.

Vazquez-Vilar, M., Quijano-Rubio, A., Fernandez-del-Carmen, A., Sarrion-Perdigones, A., Ochoa-Fernandez, R., Ziarsolo, P., Blanca, J., Granell, A., Orzaez, D., 2017. GB3.0: a platform for plant bio-design that connects functional DNA elements with associated biological data. Nucleic Acids Res. 45, 2196–2209.

Vello, V., Chu, W.-L.L., Lim, P.-E.E., Majid, N.A., Phang, S.-M.M., 2018. Metabolomic profiles of tropical Chlorella species in response to physiological changes during nitrogen deprivation. J. Appl. Phycol. 35, 1–21.

Vergunst, A.C., van Lier, M.C.M., den Dulk-Ras, A., Stüve, T. a G., Ouwehand, A., Hooykaas, P.J.J., Lier, M.C.M. Van, Dulk-ras, A. Den, Stu, T.A.G., 2005. Positive charge is an important feature of the C-terminal transport signal of the VirB/D4-translocated proteins of Agrobacterium. Proc. Natl. Acad. Sci. U. S. A. 102, 832–7.

Vigani, M., Parisi, C., Rodríguez-Cerezo, E., Barbosa, M.J., Sijtsma, L., Ploeg, M., Enzing, C., 2015. Food and feed products from micro-algae: Market opportunities and challenges for the EU. Trends Food Sci. Technol. 42, 81–92.

Vila, M., Díaz-Santos, E., De La Vega, M., Rodriǵuez, H., Vargas, Á., León, R., 2012. Promoter trapping in microalgae using the antibiotic paromomycin as selective agent. Mar. Drugs 10, 2749–2765.

Viseu, T.M.R., Hungerford, G., Coelho, A.F., Ferreira, M.I.C., 2003. Dye - Host Interactions for Local Effects Recognition in Homogeneous and Nanostructured Media. J. Phys. Chem. B 107, 13300–13312.

Viso, A.C., Marty, J.C., 1993. Fatty acids from 28 marine microalgae. Phytochemistry 34, 1521–1533.

Vogl, T., Ruth, C., Pitzer, J., Kickenweiz, T., Glieder, A., 2014. Synthetic core promoters for Pichia pastoris. ACS Synth. Biol. 3, 188–191.

Von Wirén, N., Gazzarrini, S., Gojon, A., Frommer, W.B., 2000. The molecular physiology of ammonium uptake and retrieval. Curr. Opin. Plant Biol. 3, 254–261.

Vonlanthen, S., 2013. Analysis and manipulation of storage lipids in microalgae. University College London.

Vonlanthen, S., Dauvillée, D., Purton, S., 2015. Evaluation of novel starch-deficient mutants of Chlorella sorokiniana for hyper-accumulation of lipids. Algal Res. 12, 109–118.

Walker, T.L., Collet, C., Purton, S., 2005. Algal transgenics in the genomic era. J. Phycol. 41, 1077–1093.

Wan, M., Rosenberg, J.N., Faruq, J., Betenbaugh, M.J., Xia, J., 2011. An improved colony PCR procedure for genetic screening of Chlorella and related microalgae. Biotechnol. Lett.

33, 1615–9.
Wang, C., Wang, Y., Su, Q., Gao, X., 2007. Transient expression of the GUS gene in a unicellular marine green alga, Chlorella sp.MACC/C95, via electroporation. Biotechnol. Bioprocess Eng. 12, 180–183.
Wang, J., Curtis, W.R., 2016. Proton stoichiometric imbalance during algae photosynthetic growth on various nitrogen sources: Towards metabolic pH control. J. Appl. Phycol. 28, 43–52.
Wang, S., Zhao, S.-X., Wei, C.-L., Yu, S.-Y., Shi, J.-P., Zhang, B.-G., 2014. [Effect of magnesium deficiency on photosynthetic physiology and triacylglyceride (TAG) accumulation of Chlorella vulgaris]. Huanjing kexue 35, 1462–7.
Wase, N., Tu, B., Black, P.N., DiRusso, C.C., 2015. Phenotypic screening identifies Brefeldin A/Ascotoxin as an inducer of lipid storage in the algae Chlamydomonas reinhardtii. Algal Res. 11, 74–84.
Weber, E., Engler, C., Gruetzner, R., Werner, S., Marillonnet, S., 2011. A modular cloning system for standardized assembly of multigene constructs. PLoS One 6.
Wells, M.L., Potin, P., Craigie, J.S., Raven, J.A., Merchant, S.S., Helliwell, K.E., Smith, A.G., Camire, M.E., Brawley, S.H., 2017. Algae as nutritional and functional food sources: revisiting our understanding. J. Appl. Phycol. 29, 949–982.
Werner, S., Engler, C., Weber, E., Gruetzner, R., Marillonnet, S., 2012. Fast track assembly of multigene constructs using golden gate cloning and the MoClo system. Bioeng. Bugs 3, 38–43.
Williams, P.J.L.B., Laurens, L.M.L., 2010. Microalgae as biodiesel & biomass feedstocks: Review & analysis of the biochemistry, energetics & economics. Energy Environ. Sci. 3, 554–590.
Windels, P., De Buck, S., Depicker, A., 2008. Agrobacterium tumefaciens-mediated transformation: Patterns of T-DNA integration into the host genome. In: Tzfira, T., Citovsky, V. (Eds.), Agrobacterium: From Biology to Biotechnology. Springer, New York, pp. 442–483.
Wise, A.A., Liu, Z., Binns, A.N., 2006. Culture and Maintenance of Agrobacterium Strains. In: Wang, K. (Ed.), Agrobacterium Protocols Volume 1. Human Press Inc., New Jersey, pp. 3–14.
Wodniok, S., Simon, A., Glöckner, G., Becker, B., 2007. Gain and loss of polyadenylation signals during evolution of green algae. BMC Evol. Biol. 7.
Wolf, J., Ross, I.L., Radzun, K.A., Jakob, G., Stephens, E., Hankamer, B., 2015. High-throughput screen for high performance microalgae strain selection and integrated media design. Algal Res. 11, 313–325.
Wu, C., Xiong, W., Dai, J., Wu, Q., 2015. Genome-Based Metabolic Mapping and 13C Flux Analysis Reveal Systematic Properties of an Oleaginous Microalga Chlorella protothecoides. Plant Physiol. 167, 586–99.
Wu, Y.H., Yu, Y., Hu, H.Y., 2013. Potential biomass yield per phosphorus and lipid accumulation property of seven microalgal species. Bioresour. Technol. 130, 599–602.
Xie, X., Huang, A., Gu, W., Zang, Z., Pan, G., Gao, S., He, L., Zhang, B., Niu, J., Lin, A., Wang, G., 2015. Photorespiration participates in the assimilation of acetate in Chlorella sorokiniana under high light. New Phytol. 209, 987–998.
Xing, H.-L., Dong, L., Wang, Z.-P., Zhang, H.-Y., Han, C.-Y., Liu, B., Wang, X.-C., Chen, Q.-J., 2014. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. BMC Plant Biol. 14, 327.
Xiong, W., Li, X., Xiang, J., Wu, Q., 2008. High-density fermentation of microalga Chlorella protothecoides in bioreactor for microbio-diesel production. Appl. Microbiol. Biotechnol. 78, 29–36.
Xu, Y., Boeing, W.J., 2014. Modeling maximum lipid productivity of microalgae: Review and next step. Renew. Sustain. Energy Rev. 32, 29–39.

Yamada, T., Onimatsu, H., Van Etten, J.L., 2006. Chlorella viruses. Adv. Virus Res. 66, 293–336.

Yamano, T., Iguchi, H., Fukuzawa, H., 2013. Rapid transformation of Chlamydomonas reinhardtii without cell-wall removal. J. Biosci. Bioeng. 115, 691–694.

Yanai-Balser, G.M., Duncan, G.A., Eudy, J.D., Wang, D., Li, X., Agarkova, I. V., Dunigan, D.D., Van Etten, J.L., 2010. Microarray analysis of Paramecium bursaria chlorella virus 1 transcription. J. Virol. 84, 532–42.

Yang, B., Liu, J., Jiang, Y., Chen, F., 2016. Chlorella species as hosts for genetic engineering and expression of heterologous proteins: Progress, challenge and perspective. Biotechnol. J. 11, 1244–1261.

Yang, F., Hanna, M.A., Sun, R., 2012. Value-added uses for crude glycerol - A byproduct of biodiesel production. Biotechnol. Biofuels 5, 13.

Yolken, R.H., Jones-Brando, L., Dunigan, D.D., Kannan, G., Dickerson, F., Severance, E., Sabunciyan, S., Talbot, C.C., Prandovszky, E., Gurnon, J.R., Agarkova, I. V., Leister, F., Gressitt, K.L., Chen, O., Deuber, B., Ma, F., Pletnikov, M. V., Van Etten, J.L., 2014. Chlorovirus ATCV-1 is part of the human oropharyngeal virome and is associated with changes in cognitive functions in humans and mice. Proc. Natl. Acad. Sci. U. S. A. 111, 16106–16111.

Yolken, R.H., Jones-Brando, L., Dunigan, D.D., Kannan, G., Dickerson, F., Severance, E., Sabunciyan, S., Talbot, C.C., Prandovszky, E., Gurnon, J.R., Agarkova, I. V., Leister, F., Gressitt, K.L., Chen, O., Deuber, B., Ma, F., Pletnikov, M. V., Van Etten, J.L., 2015. Reply to Kjartansdóttir et al.: Chlorovirus ATCV-1 findings not explained by contamination. Proc. Natl. Acad. Sci. U. S. A. 112, E927–E927.

Young, R.E.B., Purton, S., 2015. Codon reassignment to facilitate genetic engineering and biocontainment in the chloroplast of Chlamydomonas reinhardtii. Plant Biotechnol. J. 1–10.

YuQin, L., JinXiu, M., Di, C., Hua, X., FangXin, H., Bo, F., HongYan, Z., 2015. Proteomics analysis for enhanced lipid accumulation in oleaginous Chlorella vulgaris under a heterotrophic-Na+ induction two-step regime. Biotechnol. Lett. 37, 1021–1030.

Zhan, J., Rong, J., Wang, Q., 2017. Mixotrophic cultivation, a preferable microalgae cultivation mode for biomass/bioenergy production, and bioremediation, advances and prospect. Int. J. Hydrogen Energy 42, 8505–8517.

Zhang, J., Hao, Q., Bai, L., Xu, J., Yin, W., Song, L., Xu, L., Guo, X., Fan, C., Chen, Y., Ruan, J., Hao, S., Li, Y., Wang, R.R.-C., Hu, Z., 2014. Overexpression of the soybean transcription factor GmDof4 significantly enhances the lipid content of Chlorella ellipsoidea. Biotechnol. Biofuels 7, 128.

Zhang, P., Li, Z., Lu, L., Xiao, Y., Liu, J., Guo, J., Fang, F., 2017. Effects of stepwise nitrogen depletion on carotenoid content, fluorescence parameters and the cellular stoichiometry of Chlorella vulgaris. Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 181, 30–38.

Zhang, X.H., Tee, L.Y., Wang, X.G., Huang, Q.S., Yang, S.H., 2015. Off-target effects in CRISPR/Cas9-mediated genome engineering. Mol. Ther. - Nucleic Acids 4, e264.

Zhang, Y., Burbank, D.E., Van Etten, J.L., 1988. Chlorella viruses isolated in China. Appl. Environ. Microbiol. 54, 2170–2173.

Zhang, Y., Calin-Jageman, I., Gurnon, J.R., Choi, T., Adams, B., Nicholson, A.W., Van Etten, J.L., 2003. Characterization of a chlorella virus PBCV-1 encoded ribonuclease III. Virology 317, 73–83.

Zhang, Z., Schwartz, S., Wagner, L., Miller, W., 2000. A Greedy Algorithm for Aligning DNA Sequences. J. Comput. Biol. 7, 203–214.

Zhao, F., Yu, C., Liu, Y., 2017. Codon usage regulates protein structure and function by affecting translation elongation speed in Drosophila cells. Nucleic Acids Res. 45, 8484–8492.

Zhao, Z., Wu, X., Raj Kumar, P.K., Dong, M., Ji, G., Li, Q.Q., Liang, C., 2014. Bioinformatics Analysis of Alternative Polyadenylation in Green Alga Chlamydomonas reinhardtii Using Transcriptome Sequences from Three Different Sequencing Platforms. G3 (Bethesda). 4, 871–883.

Zheng, Y., Yu, X., Li, T., Xiong, X., Chen, S., 2014. Induction of D-xylose uptake and expression of NAD(P)H-linked xylose reductase and NADP + -linked xylitol dehydrogenase in the oleaginous microalga Chlorella sorokiniana. Biotechnol. Biofuels 7, 1–8.

Zhu, L., 2015. Biorefinery as a promising approach to promote microalgae industry: An innovative framework. Renew. Sustain. Energy Rev. 41, 1376–1384.

Zienkiewicz, K., Du, Z.Y., Ma, W., Vollheyde, K., Benning, C., 2016. Stress-induced neutral lipid biosynthesis in microalgae — Molecular, cellular and physiological insights. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids 1861, 1269–1281.

Zock, P.L., Gerritsen, J., Katan, M.B., 1996. Partial conservation of the sn-2 position of dietary triglycerides in fasting plasma lipids in humans. Eur. J. Clin. Invest. 26, 141–150.

Zuñiga, C., Li, C.-T., Huelsman, T., Levering, J., Zielinski, D.C., McConnell, B.O., Long, C.P., Knoshaug, E.P., Guarnieri, M.T., Antoniewicz, M.R., Betenbaugh, M.J., Zengler, K., 2016. Genome-Scale Metabolic Model for the Green Alga *Chlorella vulgaris* UTEX 395 Accurately Predicts Phenotypes under Autotrophic, Heterotrophic, and Mixotrophic Growth Conditions. Plant Physiol. 172, 589–602.

A APPENDIX

A.I Primers

All primers were ordered from Eurofins Genomics company.

Primer Name	Sequence (5'-3')
RBCSP-Bsa1-F	*cttGGTCTCaGGAGg*ccagaaggagcgcag
RBCSP-Bsa1-R	*cttGGTCTCaCATT*tttaagatgttgagtgacttctc
RBCST-Bsa-F	*cttGGTCTCaGCTT*gctccgtgtaaatggagg
RBCST-Bsa-R	*cttGGTCTCaAGCG*gcttcaaatacgcccagc
CaMVP-Bsa-F(R)	*cttGGTCTCaCATT*agagatagatttgtagagagagactg
CaMVP-Bsa-R(F)	*cttGGTCTCaGGAG*atggtggagcacgac
CaMVT-Bsa-F(R)	*cttGGTCTCaAGCG*taattcgggggatctgg
CaMVT-Bsa-R(F)	*cttGGTCTCcGCTT*gtcgatcgacaagctc
mChBsa_f	*gtatGGTCTCa*AATGGTGTCCAAGGG
tdToBsa_f	*ggtatGGTCTCa*AATGACCAGCAAGG
mCh-tdTo-Bsa_r	*taatGGTCTCt*AAGCTTACTTGTACAGCTCGT
Ble2I-Bsa-F	*cttGGTCTCt*AATGGCCAGgtgagtc
Ble2I-Bsa-R	*cttGGTCTCa*AAGCTTAGTCCTGCTCCTCG
BLE-Bsa-F	*cttGGTCTCt*AATGGCCAAGCTGAC
BLE-Bsa-R	*cttGGTCTCt*AAGCTTAGTCCTGCTCCTCG
Od6Bsa_f	*gtatGGTCTCt*AATGTGTGTTGAGACGGAGAAC
Od6Bsa_r	*taatGGTCTCt*AAGCTCAAGCGGTCTTTCCAG
Od6-domF	ggctcgccggtagcg
Od6-domR	cgaacccggctgtgaag
J23RFP-*BsaI*-div-F	*ct*GGAG*aGAGACC*ttgacggctagctcag
J23RFP-*BsaI*-div-R	*gga*AGCG*tGAGACC*tctagtatataaacgcagaaagg
pCAMpVS1Bsa-fwd	GGAGTACATCGAGATCGAGCTAG
pCAMpVS1Bsa-rev	TATGGTCTCAGGAAACAGACCTTTTCG

Table A-1 Primers used to generate the library of Level 0 parts and Level 1 acceptor plasmid. Where present, the 4 bp syntax assembly sequence is underlined. The *BsaI* site is GGTCTC and is in italic capitals. Any portions of the primer which overhang from the template sequence are in italics. Annealing temperatures were determined from the manufacturer recommendations for the specific polymerases used (New England Biolabs Phusion® or Q5 ® or Taq)

Primer Name	Description	Sequence (5'-3')
pJET1.2-scr-F pJET1.2-scr-R	Used to confirm sequence of PCR-generated Level 0 parts in housing vector pJET 1.2 (Thermo). Must add 303 bp on to expected product size.	F: TTTTAACTTGGAGCAGGTTCC R: GTTTTCATGAGAGTCGATTGC
pC23-scr F pC23-scr-R	For colony PCR or sequencing of Golden-Gate assembled PCR products into the binary vector pC23c via pC23c-RFP. Must add 254 bp to expected product size.	F: gctttacactttatgcttccg R: cttcgctattacgccagc
J23-mid-R	Reverse primer for use with pC23-scr-F when size of colony PCR product is too similar to the J23100 cloning insert in pC23c-RFP to determine if it correct. This primer binds inside the RFP gene, hence it should not produce a product if the construct is correct.	R: gaagccatctagtatttctcctc
pTiBo542-F pTiBo542-R	For testing presence of *Agrobacterium* Ti plasmid for EHA105. Primer sequence is from Deeba *et al.* (Deeba et al., 2014)	F: CCCGCTGAGAATGACGCCAA R: CCTGCGACACATCGTTGCTGA
ITS1 ITS4	Positive control primers for algae genomic preps. Amplifies internal spacer region ITS-2 of ribosome. Sequences from (Hall et al., 2010).	ITS1: AGGAGAAGTCGTAACAAGGT ITS4: TCCTCCGCTTATTGATATGC

Table A-2 Sequencing primers used for bacterial and algal clones. Used for confirmation of clones or for sequencing.

A.II Survey of published protocols for *Agrobacterium*-mediated transformation of microalgae

Table A-3 Summary of published protocol optimisations for *Agrobacterium*-mediated transformation of microalgae. Notes are shown on efficiency and parameters optimised.

Algal Species	Type	Reference	Parameter(s) optimised	Range tested	Optimal value	Transformation efficiency improvement	Notes
Chlamydomonas reinhardtii CC-124 (mt-)	FW	(Kumar et al., 2004; Rajam and Kumar, 2007)	Acetosyringone	0 or 100 µM	100 µM	~50×	
			Co-incubation media	Solid or liquid	Solid	"greatly increased"	
Chlamydomonas reinhardtii CC-125 (mt+)	FW	(Pratheesh et al., 2012)	*Agrobacterium* strain	EHA101, EHA105, LBA4404	EHA105	~2×	No AS
		(Pratheesh et al., 2014)	Induction step before co-incubation: pH 5.2, AS 100 µM, 1 mM glycine betaine. 25 °C, 4h with addition of algae for 30 min before plating on solid media	Presence or absence	Presence	"Double"	Direct comparison to Kumar et al., 2004
		(Mini et al., 2018)	*Agrobacterium* strain	C58C1, LBA4404	C58C1	~1.2×	
			Agrobacterium virulence using acidic AB medium for *Agrobacterium* pre-culture and co-cultivation	Presence or absence	Only successful with presence		
Chlamydomonas reinhardtii cw15	FW	(Mini et al., 2018)	*Agrobacterium* strain	C58C1, LBA4404	Only successful with C58C1		
			Agrobacterium virulence, see above				
Chlorella vulgaris (Jallo Park, Lahore, Pakistan)	FW	(Sharif et al., 2015)	Co-cultivation length	2-5 days	3 days	~2×	
			Co-cultivation temperature	20-30 °C	25 °C	~2×	
			Co-cultivation medium pH	pH 5.0-6.5	pH 5.5	~2×	
			Density of *Agrobacterium* in co-culture	OD 600 0.5-1.5	OD 600 1.0	~2×	
			Acetosyringone concentration in co-cultivation media (solid)	50-150 µM	100 µM	~3×	
Chlorella vulgaris CGMCC 6951	FW	(Lou et al., 2016)	*Agrobacterium* strain	EHA105, LBA4404	EHA105	Visibly many more selective colonies obtained	Protoplasts used
Chlorella vulgaris UMT-M1	FW	(Cha et al., 2012, 2011b)	Pre-culture algae duration (solid)	0-5 days	2 days	>65× ($p<0.05$)	Combined optimised parameter average 25.0% GUS-positive pre-selection
			Co-cultivation length	1-5 days	3-4 days	~5× ($p<0.05$)	
			Co-cultivation temperature	20-30 °C	24-25 °C	~12× ($p<0.05$)	
			Co-cultivation medium pH	pH 5.0-6.0	5.5-5.6	~2× ($p<0.05$)	

			Density of *Agrobacterium* in co-culture	OD 600 0.2-1.0	OD 600 1.0	~2×($p<0.05$)	
			Acetosyringone concentration in co-cultivation media (solid)	0-300 µM	≥150 µM	~1.5×($p<0.05$); no transformants at 0 µM	
Dunaliella bardawil V-101	M	(Anila et al., 2011)	Acetosyringone concentration	0-200 µM	n/a	No effect observed	Cells adapted to low salt (0.2 M)
Dunaliella salina V-101	M	(Srinivasan and Gothandam, 2016)	*Agrobacterium* strain	EHA105, GV3101, LBA4404	GV3101	Up to 2×	Cells adapted to low salt (0.15 M). LBA4404 highest efficiency but double insertion instead of single.
			Acetosyringone and D-glucose concentration in *Agrobacterium* induction media at pH 5.2	AS: 0 or 100 µM; D-glc: 5-15 mM; Combinations	10 mM D-glc + 100 µM AS	Up to 4×. No effect if not combined.	
Haematococcus pluvialis SAG 19-a	FW	(Kathiresan et al., 2009) 2009a	AS concentration in co-cultivation media	0-250 µM	100 µM	~1.4×	AS not essential.
Haematococcus pluvialis SAG 19-a	FW	(Kathiresan and Sarada, 2009) 2009b	Co-cultivation media	BBM+1/2 LB, Z8+1/2 LB, Z8+0.5% mannitol, Z8, TAP	TAP	Only TAP supported both *Agrobacterium* and algae growth.	No AS
Isochrysis galbana KMMCC-12 and *Isochrysis* sp. H-13	M	(Prasad et al., 2014)	Co-cultivation duration	24-72 hours	KM: 48 h H-13: 72 h	Comparing all optimised to all non-optimised: KM: ~100× H-13: ~10×	
			Co-cultivation medium	ASW, ASW+YEB, ASW+IM	KM: ASW H-13: ASW		
			Acetosyringone concentration	100-300 µM	KM: 100 µM H-13: 200 µM		
Nannochloropsis sp.	M	(Cha et al., 2011b)	Pre-culture algae duration (solid)	0-5 days	5 days	~53×	Combined optimised parameter average 24.6% GUS-positive pre-selection
			Co-cultivation length	1-5 days	3-5 days	~7× ($p<0.05$)	
			Co-cultivation temperature	20-30 °C	24-25 °C	~48× ($p<0.05$)	
			Co-cultivation medium pH	pH 5.0-6.0	pH 5.5-5.6	~5× ($p<0.05$)	
			Density of *Agrobacterium* in co-culture	OD 600 0.2-1.0	OD 600 1.0	~3× ($p<0.05$)	
			Acetosyringone concentration in co-cultivation media	0-300 µM	50-100 µM	~8× ($p<0.05$)	
Nannochloropsis sp. UMT-M3	M	(Cha et al., 2011a)	Phenolic inducer compound	Acetosyringone, Vanillin, Cinnamic Acid, Coumarin	Cinnamic acid at 200-400 µM	Up to 10× ($p<0.05$)	*Agrobacterium* strain LBA4404
			Concentration of phenolic inducer	0-600 µM			
Penium margaritaceum	FW	(Sorensen et al., 2014)	Pre-culture medium	MI medium, WHM medium	MI medium	n/s	MI medium is Nitrogen starved and induces conjugation life-stage of *Penium* cells.
			Pre-culturing duration	1-5 days	3-5 days	n/s; “greatly enhanced”	
			Co-cultivation duration	1-5 days	1 day	n/s	
			Acetosyringone concentration	0-500 µM	500 µM	n/s; “highest number initial transgenic cells”	
			Co-cultivation medium	Solid or Liquid	Solid	Unsuccessful in liquid	
	FW		Algae pre-culture media pH	pH 5.5 or 6.7	5.5	Suggested optimal value	

Scenedesmus almeriensis CCAP 276/24		(Dautor et al., 2014)	Density of *Agrobacterium* in co culture	OD 600 0.5 or 1.0	OD 600 0.5	Suggested optimal value	The suggested optimal values were not statistically significant differences but were suggested as a set of optimised values by the statistics software.
			Washing of algae cells with induction medium (150 µM AS) before co-culturing	Presence or absence of step	Presence	Suggested optimal value	
			Co-culturing duration	2 or 3 days	2 days	Suggested optimal value	
			Co-culturing light	Presence or absence	No effect	No effect	
			Co-culturing temperature	22 or 26 °C	22	Significant	
			Post-cocultivation algae recovery period containing no selection apart from *Agrobacterium* killing	Presence or absence of step	Absence	Suggested optimal value	
Schizochytrium sp. TIO1101 (CGMCC #4603)	M	(Cheng et al., 2012)	*Agrobacterium* strain	LBA4404, EHA105	LBA4404	"more transformants"	Protoplasts used
			Acetosyringone concentration in *Agrobacterium* induction step	0 or 200 µM	200 µM	AS is essential for success	
Tetraselmis chuii CCAP 66/21B	M	(Ubeda-Minguez et al., 2015)	Acetosyringone concentration in algae pre-culture wash and co-incubation plate	100-300 µM	150 µM	Suggested optimal value	First-order interaction between AS concentration and co-culturing length. pH and temperature showed quadratic response. The suggested optimal values were not statistically significant differences but were suggested as a set of optimised values by the statistics software.
			pH of media in algae pre-culture wash and co-incubation plate	pH 5.0 to 6.7	pH 5.0	Significant	
			Agrobacterium density at co-incubation	OD 600 0.5 or 1.0	OD 600 1.0	Suggested optimal value	
			Co-culture length	2 or 3 days	3 days	Suggested optimal value	
			Co-culture temperature	22-30 °C	27 °C	Significant	
			Co-culture light	Presence or absence	No effect	No effect	
			Post-cocultivation algae recovery period containing no selection apart from *Agrobacterium* killing	Presence or absence of step	Absence	Suggested optimal value	

Table A-4 Summary of some key parameters used in protocols for Agrobacterium-mediated transformation of microalgae. Parameters which were optimised by the publication are underlined. AS: Acetosyringone, PEG: polyethylene glycol, TAP: Tris-Acetate Phosphate medium.

Algal Species	**Type**	**Publication**	***Agrobacterium* strain**	***Agrobacterium* induction**			**Algae pre-culture**	**Co-incubation**			**Binary vector backbone**
				Chemical and concentration	**pH**	**When induction was carried out**		**Media**	**Temp**	**Time**	
Ankistrodesmus braunii (Naegeli) Brunnthaler strain ACUF 160-146	FW	(Chiaiese et al., 2011)	LBA4404	n/s, protocol specified as Kumar et al., 2004 except use of BBM media.							pGreen 0029
Ankistrodesmus sp. JX456463 (RS-2012)	FW	(Sanitha et al., 2014)	LBA4404	None	7.0	None	Solid	Solid	n/s, 25 °C	48 h	pCAMBIA 1301
Chlamydomonas pitschmannii Ettl. ACUF 238/79	FW	(Chiaiese et al., 2011)	LBA4404	n/s, protocol specified as Kumar et al., 2004 except use of BBM media							pGreen 0029
Chlamydomonas reinhardtii CC-124 (mt-)	FW	(Kumar et al., 2004; Rajam and Kumar, 2007)	LBA4404	AS, <u>100 µM</u>	n/s. TAP media.	AS in algae pre-incubation/co-cultivation plate, also added to *Agrobacterium* upon plating for co-incubation	Solid	<u>Solid</u>	n/s, 23 °C	48 h	pCAMBIA 1304
		(Hema et al., 2007)	EHA101	AS, 100 µM	n/s. TAP media.	AS in algae pre-incubation/co-cultivation plate	Solid	Solid	n/s, 23 °C	48 h	pCAMBIA 1301; pGAH
Chlamydomonas reinhardtii CC-125 (mt+)	FW	(Pratheesh et al., 2012)	<u>EHA105</u>	None	n/s. TAP media	None	Solid	Solid	25 °C	2 d	pCAMBIA 1304
		(Pratheesh et al., 2014)	EHA105	AS, 100 µM; Glycine betaine, 1 mM	5.2	<u>Separate step 4 h before co-cultivation including 30 min with algae</u>	Liquid	Solid	25 °C	2 d	pCAMBIA 1301
		(Mini et al., 2018)	C58C1	AS, 100 µM; 0.5 % glucose	5.6	Overnight induction of *Agrobacterium*	Liquid	Solid	n/s	48 h	pCAMBIA 1390
Chlamydomonas reinhardtii cw15	FW	(Mini et al., 2018)	As above								
Chlamydomonas reinhardtii CC-137 (mt+)	FW	(Soria-Guerra et al., 2014)	GV3101	AS, 100 µM	n/s. TAP media	AS in algae pre-incubation/co-cultivation plate	Solid	Solid	n/s, 25 °C	48 h	pBI-121
Chlorella emersonii Shihira and Krauss. ACUF 061-103, 317-24 and 053-95	FW	(Chiaiese et al., 2011)	LBA4404	n/s, protocol specified as Kumar et al., 2004 except use of BBM media.							pGreen 0029
Chlorella pyrenoidosa NCIM, India	FW	(Reddy et al., 2017)	LBA4404	AS, 100 µM	n/s. ACB media	AS added to *Agrobacterium* just before plating for co-incubation	Solid	Solid	25 °C	3 d	pART27

Chlorella sp. (Iso 5 GenBank JX041601)	FW	(Sanitha et al., 2014)	LBA4404	None	7.0	None	Solid	Solid	n/s, 25 °C	48 h	pCAMBIA 1301
Chlorella sp. DT	FW	(Lin et al., 2013)	LBA4404	None	n/s	None	Liquid	Liquid	28 °C	2 d	pHm3A
Chlorella vulgaris (Jallo Park, Lahore, Pakistan)	FW	(Sharif et al., 2015)	EHA101	AS, 100 µM	5.5	Algae pre-incubation/co-cultivation plate and separate step *Agrobacterium* 30 min before co-incubation	Solid	Solid	25 °C	3 d	n/s
Chlorella vulgaris CGMCC 6951	FW	(Ma and Lin, 2014)	LBA4404	AS, 150 µM; 10 mM glucose; 0.5% (w/v) glycerol	5.3	Separate step 4-5 h before co-cultivation	Protoplasts made, liquid	Liquid	28 °C	24 h	pCAMBIA 2301
		(Lou et al., 2016)	EHA105	AS, 200 µM; 10 mM glucose; 0.5% (w/v) glycerol	5.3	Separate step 4-5 h before co-cultivation	Protoplasts made, liquid	Liquid	28 °C	24	pCAMBIA 1302
Chlorella vulgaris UMT-M1	FW	(Cha et al., 2012, 2011b)	LBA4404	AS, 150 µM	5.5	*Agrobacterium* and algae pre-culture wash, co-cultivation media.	Solid	Solid	24 °C	3 d	pCAMBIA 1304
		(Lau et al., 2017)	LBA4404	AS, 50 µg/ml	n/s	*Agrobacterium* pre-culture wash, co-cultivation medium	Liquid, cells treated with enzymes	Solid	27 °C	3 d	pCAMBIA 1304
		(Ng et al., 2016)	LBA4404	AS, 100 µM	5.6	Algae pre-culture wash, co-incubation media.	Liquid, enzyme pre-treatment	Solid	n/s	3 d	pMDC150/ pMDC221
Dunaliella bardawil ATCC #30861	M	(Fang et al., 2012)	GV3101	AS, 100 µM	n/s	Co-incubation media	Liquid, adaptation to low salt	Liquid	n/s	2 d	pGreen 0229-35S
Dunaliella bardawil V-101 (Madras University, Chennai)	M	(Anila et al., 2011)	EHA101	None needed	n/s; TAP medium	none	Solid, after adaptation to low salt	Solid	22 °C	48 h	pCAMBIA 1304
Dunaliella salina V-101	M	(Anila et al., 2015)	EHA101	n/s, protocol specified as Anila et al., 2011.							pCAMIBA 1304
Dunaliella salina V-101	M	(Srinivasan and Gothandam, 2016)	GV3101	AS, 100 µM, and D-glucose 10 mM	5.2	Separate step 24 h before co-cultivation	Solid, after adaptation to low salt	Solid	n/s	48 h	pMDC45
Dunaliella salina (Madras University, Chennai)	M	(Simon et al., 2016)	EHA 101	n/s, protocol specified as Anila et al., 2011.							pCAMBIA 1304
Haematococcus pluvialis SAG 19-a	FW	(Kathiresan et al., 2009)	EHA 101	AS, 100 µM	n/s; TAP medium	Co-incubation media	Solid	Solid	22 °C	48 h	pCAMBIA 1301
Haematococcus pluvialis SAG 19-a	FW	(Kathiresan and Sarada, 2009)	EHA 101	None	n/s; TAP medium	None	Solid	Solid	22 °C	48 h	pCAMBIA 1301
Haematococcus pluvialis SAG 34-1a	FW	(Kathiresan et al., 2015)	EHA101	None	n/s; TAP medium	None	Solid	Solid	22 °C	48 h	pCAMBIA 1304

Isochrysis galbana KMMCC-12 and *Isochrysis* sp. H-13	M	(Prasad et al., 2014)	LBA4404	AS, 100 -200 µM	8.0	Co-incubation media	Liquid	Liquid	25 °C	48-72 h	pCAMBIA 1380
Nannochloropsis sp.	M	(Cha et al., 2011b)	LBA4404	AS, 50 µM	5.5	Algal pre-culture/co-incubation plate	Solid	Solid	24 °C	3 d	pCAMBIA 1304
Nannochloropsis sp. UMT-M3	M	(Cha et al., 2011a)	LBA4404	Cinnamic acid 200-400 µM	5.6	Co-incubation plate, pre-culture wash (pH only)	Liquid	Solid	25 °C	3 d	pCAMBIA 1304
ParaChlorella kessleri isolated from the Bay of Bengal, India	M	(Rathod et al., 2013)	LBA4404	None	n/s; TAP medium	None	Solid	Solid	25 °C	72 h	pCAMBIA 1301
Penium margaritaceum	FW	(Sorensen et al., 2014)	GV2260	AS, 500 µM	7.5	Separate step 3 h before incubation	Liquid	Solid	Room temp.	24 h	pCAMBIA 1302 & 1301, pART27
Scenedesmus almeriensis CCAP 276/24	FW	(Dautor et al., 2014)	LBA4404	AS, 150 µM	5.5	Cell resuspension before plating for co-cultivation at specified pH.	Solid	Solid	22 °C	2 d	pCAMBIA 1305.1
Scenedesmus bajacalifornicus JQ782744	FW	(Sanitha et al., 2014)	LBA4404	None	7.0	None	Solid	Solid	n/s, 25 °C	48 h	pCAMBIA 1301
Schizochytrium sp. TIO1101 (CGMCC #4603)	M	(Cheng et al., 2012)	LBA4404	AS, 200 µM	n/s	Separate step 4 h before co-cultivation	Protoplasts made, liquid	Liquid	28 °C	12 h	pCAMBIA 2301
Symbiodinium kawagutii Trench & Blank, *Symbiodinium sp.* Mf11.5b.1, *Symbiodinium microadriaticum* subsp. microadriaticum (MAC-CassKB8)	M, E	(Ortiz-Matamoros et al., 2015)	GV3101	None	n/s	None	Liquid, abrasion with glass beads & PEG.	Liquid	n/s, 25 °C	1-2 d	pCB302
Tetraselmis chuii (or *chui*) CCAP 66/21B	M	(Ubeda-Minguez et al., 2015)	LBA4404	AS, 150 µM	5.0	Algae pre-culture wash, co-incubation plate	Solid, half strength salinity	Solid	27 °C	3 d	pCAMBIA Shble
		(Úbeda-Mínguez et al., 2017)	Protocol stated as that of Ubeda-Minguez et al., 2015								pCAMBIA Shble

Table A-5 Summary of transformation efficiencies, integration pattern and transgene stability for published reports of *Agrobacterium*-mediated transformation of microalgae. CFU: Colony Forming Units.

Algal Species	Publication	Maximum transformation frequency	Frequency calculation method	Transformant stability	Integration pattern
Ankistrodesmus braunii (Naegeli) Brunnthaler strain ACUF 160-146	(Chiaiese et al., 2011)	7.76×10^{-6}	Colony frequency after 20 days incubation in selective conditions	n/s	n/s
Ankistrodesmus sp. JX456463 (RS-20120)	(Sanitha et al., 2014)	3.50 % (no inducer)	GUS assay after hygromycin selection and colony PCR	n/s	n/s
Chlamydomonas pitschmannii Ettl. ACUF 238/79	(Chiaiese et al., 2011)	1.16×10^{-6} (inducer not stated)	Colony frequency after 20 days incubation in selective conditions	n/s	n/s
Chlamydomonas reinhardtii CC-124 (mt-)	(Kumar et al., 2004)	11 % (GUS) $311\text{-}355 \times 10^{-6}$ (Hyg^R)	GUS assay after coincubation and Hyg^R colony frequency	18 months antibiotic free medium	Random. Single and multiple (mainly double) integration events observed
	(Hema et al., 2007)	$7\text{-}8 \times 10^{-6}$	GUS lines, PCR screen	n/s	Random. Single and double insertions.
	(Kumar and Rajam, 2007)	8-10 % (GUS) $5\text{-}8 \times 10^{-6}$ (Hyg^R)	GUS or Hyg^R colony frequency	GUS transient. Hyg stable.	n/s
	(Rajam and Kumar, 2007)	$311\text{-}355 \times 10^{-6}$	Hyg^R colony frequency	n/s	n/s
Chlamydomonas reinhardtii CC-125 (mt+)	(Pratheesh et al., 2012)	$13.82 \pm 2.8 \times 10^{-6}$ (no inducer)	Hyg^R colony frequency	n/s	n/s
	(Pratheesh et al., 2014)	$523 \pm 10.2 \times 10^{-6}$	Hyg^R colony frequency	n/s	Random, single copy (n=1)
	(Mini et al., 2018)	$0\text{-}14 \times 10^{-8}$ of which 0-92 % are PCR positive.	$Paro^R$ colony frequency	Luciferase activity tested for 20 subcultures and was stable.	Random, 1-2 copies (mainly 1)
Chlamydomonas reinhardtii CC-137 (mt+)	(Soria-Guerra et al., 2014)	n/s; "low". The authors note potential toxicity of the foreign protein	Kan^R selection	n/s	Copy number single and double via RT-PCR (67% single 33% double, n=6)
Chlamydomonas reinhardtii cw15	(Mini et al., 2018)	$16\text{-}33 \times 10^{-8}$ of which 67-92 % are PCR positive.	$Paro^R$ colony frequency	Luciferase activity tested for 20 subcultures and was stable.	Random, 1-2 copies (mainly 1)
Chlorella emersonii Shihira and Krauss. ACUF 061-103, 317-24 and 053-95	(Chiaiese et al., 2011)	9.36×10^{-6} (#103) 8.5×10^{-6} (#24) 24.3×10^{-6} (#95)	Colony frequency after 20 days incubation in selective conditions	n/s	n/s
Chlorella pyrenoidosa NCIM, India	(Reddy et al., 2017)	n/s	n/s	n/s	n/s
Chlorella sp. (Iso 5 GenBank JX041601)	(Sanitha et al., 2014)	12.25%	GUS assay after hygromycin selection and colony PCR	n/s	n/s
Chlorella sp. DT	(Lin et al., 2013)	n/s	n/s	n/s	n/s

Chlorella vulgaris (Jallo Park, Lahore, Pakistan)	(Sharif et al., 2015)	12-16%	Transient GUS assay after co-cultivation (no selection)	n/s	n/s
Chlorella vulgaris CGMCC 6951	(Ma and Lin, 2014)	n/s; "hundreds of colonies"	G418 selection	n/s	n/s
	(Lou et al., 2016)	n/s; "numerous clones"	Hyg selection	n/s	n/s
Chlorella vulgaris UMT-M1	(Cha et al., 2012, 2011b)	25 % (GUS) 33 % (Hyg^R colony PCR) of which 90% are PCR positive for GUS.	Transient GUS assay 2 days after co-cultivation (no selection) or PCR for Hyg and GUS of Hyg^R colonies after 20 days selection.	Loss of GUS expression but maintenance of Hyg^R in the lines showing PCR positive for Hyg and GUS	n/s
	(Lau et al., 2017)	19%	PCR for transgene in Hyg^R selected clones	Phenotype reverted after 6 months	n/s
	(Ng et al., 2016)	n/s	n/a	>1 year	n/s
Dunaliella bardawil ATCC #30861	(Fang et al., 2012)	n/s; a single transformant is analysed	n/s	n/s	n/s
Dunaliella bardawil V-101 (Madras University, Chennai)	(Anila et al., 2011)	$42\pm3 \times 10^{-6}$	Hyg^R CFU after 8 weeks	18 months on antibiotic free plates	n/s
Dunaliella salina V-101	(Anila et al., 2015)	n/s	n/s	n/s	Multiple random integrations, mainly 2
	(Srinivasan and Gothandam, 2016)	$61\text{-}181 \times 10^{-6}$ of which 81-95 % show bands on southern blot, both depending on *Agrobacterium* strain.	Hyg^R CFU frequency and Southern blot	>6 months phenotype and PCR positive	Random, 1-2 copies depending on *Agrobacterium* strain
Dunaliela salina (Madras University, Chennai)	(Simon et al., 2016)	$40\pm5 \times 10^{-6}$	Hyg^R CFU frequency	6 months but reduced phenotype for 9/10 tested.	Random, 2 copies
Haematococcus pluvialis SAG 19-a	(Kathiresan et al., 2009)	$109\text{-}153 \times 10^{-6}$ depending on inducer concentration	Hyg^R CFU frequency	2.5 years non-selective medium and positive PCR and GFP.	Random, 2-3 copies, though authors note this could be due to mixture of cells or incomplete restriction enzyme digestion.
	(Kathiresan and Sarada, 2009)	109.64×10^{-6}	Hyg^R CFU frequency	1.5 years on non-selective medium	n/s
Haematococcus pluvialis SAG 34-1a	(Kathiresan et al., 2015)	n/s	n/s	n/s	Multiple (< 3 copies), random
Isochrysis galbana KMMCC-12 and *Isochrysis* sp. H-13	(Prasad et al., 2014)	58×10^{-5} (KMMCC-12) 850×10^{-5} (H-13)	Norflurazon resistant CFU compared to no selection.	24 months	Random, 1-3 copies, mainly 1
Nannochloropsis sp.	(Cha et al., 2011b)	24.55 %	Transient GUS assay 2 days post co-cultivation (no selection)	n/s	n/s
Nannochloropsis sp. UMT-M3	(Cha et al., 2011a)	21 % (novel inducer)	Transient GUS assay 2 days post co-cultivation (no selection)	n/s	n/s
ParaChlorella kessleri (isolate from Bay of Bengal, India)	(Rathod et al., 2013)	$246\text{-}251 \times 10^{-7}$	Frequency: Hyg^R CFU per ml	Four generations	n/s
Penium margaritaceum	(Sorensen et al., 2014)	n/s	n/s	Multiple subcultures	n/s
Scenedesmus almeriensis CCAP 276/24	(Dautor et al., 2014)	1×10^{-4} (Hyg), of which 80% PCR +ve	Hyg^R or Zeo^R colonies per initial microalgal cell and PCR screen	24 selective subcultures (Hyg)	n/s

		0.3-0.9 × 10^{-4} (Zeo), of which 50% PCR +ve		30 % stable (12 month) survival after 2 selective subcultures (Zeo)	
Scenedesmus bajacalifornicus JQ782744	(Sanitha et al., 2014)	2.96 %	GUS assay after hygromycin selection and colony PCR	n/s	n/s
Schizochytrium sp. TIO1101 (CGMCC #4603)	(Cheng et al., 2012)	n/s; 60-170 transformants depending on *Agrobacteirum* strain.	G418 selection	n/s	Random single copies as detected for eGFP southern
Symbiodinium sp.: *Symbiodinium kawagutii* Trench & Blank, *Symbiodinium sp.* Mf11.5b.1, *Symbiodinium microadriaticum* subsp. microadriaticum (MAC-CassKB8)	(Ortiz-Matamoros et al., 2015)	790-839 × 10^{-6} (kawagutii) 421-460 × 10^{-6} (Mf11.) 570-640 × 10^{-6} (KB8)	GFP fluorescence detection frequency after selection	After selection cells did not divide further	n/s
Tetraselmis chuii CCAP 66/21B	(Ubeda-Minguez et al., 2015)	0.84-2.16 × 10^{-4}	Ratio of Phl^R CFU to initial number of cells	6 months non-selective were Phl^R of which 60% were GUS +ve	n/s
	(Úbeda-Mínguez et al., 2017)	n/s “hundreds of putative transformants”	Phl^R CFU	From hundreds of putative transformants, ~50 were stable phenotype	n/s

A.III Alignment of CaMV35S3 from pCAMBIA-2300 with other sources

```
CLUSTAL O(1.2.4) multiple sequence alignment

GB0222_2x35S_annotated(fwd)          actagagccaagctgatctcctttgccccggagatcaccatggacgactttctctatctc   60
pGWB602omega_2x35S_annotated(fwd)    ------------------------------------------------------------   0
pCAMBIA2300_2x35S_annoted(fwd)       ------------------------------------------------------------   0
pCAMBIA1304_2x35S_annotated(fwd)     ------------------------------------------------------------   0

GB0222_2x35S_annotated(fwd)          tacgatctaggaagaaagttcgacggagaaggtgacgataccatgttcaccaccgataat   120
pGWB602omega_2x35S_annotated(fwd)    ------------------------------------------------------------   0
pCAMBIA2300_2x35S_annoted(fwd)       ------------------------------------------------------------   0
pCAMBIA1304_2x35S_annotated(fwd)     ------------------------------------------------------------   0

GB0222_2x35S_annotated(fwd)          gagaagattagcctcttcaatttcagaaagaatgctgacccacagatggttagagaggcc   180
pGWB602omega_2x35S_annotated(fwd)    ------------------------------------------------------------   0
pCAMBIA2300_2x35S_annoted(fwd)       ------------------------------------------------------------   0
pCAMBIA1304_2x35S_annotated(fwd)     ------------------------------------------------------------   0

GB0222_2x35S_annotated(fwd)          tacgcggcaggtctgatcaagacgatctacccgagtaataatctccaggagatcaaatac   240
pGWB602omega_2x35S_annotated(fwd)    ------------------------------------------------------------   0
pCAMBIA2300_2x35S_annoted(fwd)       ------------------------------------------------------------   0
pCAMBIA1304_2x35S_annotated(fwd)     ------------------------------------------------------------   0

GB0222_2x35S_annotated(fwd)          cttcccaagaaggttaaagatgcagtcaaaagattcaggactaactgcatcaagaacaca   300
pGWB602omega_2x35S_annotated(fwd)    ------------------------------------------------------------   0
pCAMBIA2300_2x35S_annoted(fwd)       ------------------------------------------------------------   0
pCAMBIA1304_2x35S_annotated(fwd)     ------------------------------------------------------------   0

GB0222_2x35S_annotated(fwd)          gagaaagatatatttctcaagatcagaagtactattccagtatggacgattcaaggcttg   360
pGWB602omega_2x35S_annotated(fwd)    ------------------------------------------------------------   0
pCAMBIA2300_2x35S_annoted(fwd)       ------------------------------------------------------------   0
pCAMBIA1304_2x35S_annotated(fwd)     ------------------------------------------------------------   0

GB0222_2x35S_annotated(fwd)          cttcataaaccaaggcaagtaatagagattggagtctctaagaaagtagttcctactgaa   420
pGWB602omega_2x35S_annotated(fwd)    ------------------------------------------------------------   0
pCAMBIA2300_2x35S_annoted(fwd)       ------------------------------------------------------------   0
pCAMBIA1304_2x35S_annotated(fwd)     ------------------------------------------------------------   0

GB0222_2x35S_annotated(fwd)          tcaaaggccatggagtcaaaaattcagatcgaggatctaacagaactcgccgtgaagact   480
pGWB602omega_2x35S_annotated(fwd)    ------------------------------------------------------------   0
pCAMBIA2300_2x35S_annoted(fwd)       ------------------------------------------------------------   0
pCAMBIA1304_2x35S_annotated(fwd)     ------------------------------------------------------------   0

GB0222_2x35S_annotated(fwd)          ggcgaacagttcatacagagtcttttacgactcaatgacaagaagaaaatcttcgtcaac   540
pGWB602omega_2x35S_annotated(fwd)    ------------------------------------------------------------caac   4
pCAMBIA2300_2x35S_annoted(fwd)       ------------------------------------------------------------   0
pCAMBIA1304_2x35S_annotated(fwd)     ----------------------------------gcgtattggctagagcagcttgccaac   27

GB0222_2x35S_annotated(fwd)          atggtggagcacgacactctcgtctactccaagaatatcaaagatacagtctcagaagac   600
pGWB602omega_2x35S_annotated(fwd)    atggtggagcacgacactctcgtctactccaagaatatcaaagatacagtctcagaagac   64
pCAMBIA2300_2x35S_annoted(fwd)       atggtggagcacgacactctcgtctactccaagaatatcaaagatacagtctcagaagac   60
pCAMBIA1304_2x35S_annotated(fwd)     atggtggagcacgacactctcgtctactccaagaatatcaaagatacagtctcagaagac   87
                                     ************************************************************

GB0222_2x35S_annotated(fwd)          caaagggctattgagacttttcaacaaagggtaatatcgggaaacctcctcggattccat   660
pGWB602omega_2x35S_annotated(fwd)    caaagggctattgagacttttcaacaaagggtaatatcgggaaacctcctcggattccat   124
pCAMBIA2300_2x35S_annoted(fwd)       caaagggctattgagacttttcaacaaagggtaatatcgggaaacctcctcggattccat   120
pCAMBIA1304_2x35S_annotated(fwd)     caaagggctattgagacttttcaacaaagggtaatatcgggaaacctcctcggattccat   147
                                     ************************************************************

GB0222_2x35S_annotated(fwd)          tgcccagctatctgtcacttcatcaaaaggacagtagaaaaggaaggtggcacctacaaa   720
pGWB602omega_2x35S_annotated(fwd)    tgcccagctatctgtcacttcatcaaaaggacagtagaaaaggaaggtggcacctacaaa   184
pCAMBIA2300_2x35S_annoted(fwd)       tgcccagctatctgtcacttcatcaaaaggacagtagaaaaggaaggtggcacctacaaa   180
pCAMBIA1304_2x35S_annotated(fwd)     tgcccagctatctgtcacttcatcaaaaggacagtagaaaaggaaggtggcacctacaaa   207
                                     ************************************************************

GB0222_2x35S_annotated(fwd)          tgccatcattgcgataaaggaaaggctatcgttcaagatgcctctgccgacagtggtccc   780
pGWB602omega_2x35S_annotated(fwd)    tgccatcattgcgataaaggaaaggctatcgttcaagatgcctctgccgacagtggtccc   244
pCAMBIA2300_2x35S_annoted(fwd)       tgccatcattgcgataaaggaaaggctatcgttcaagatgcctctgccgacagtggtccc   240
pCAMBIA1304_2x35S_annotated(fwd)     tgccatcattgcgataaaggaaaggctatcgttcaagatgcctctgccgacagtggtccc   267
                                     ************************************************************

GB0222_2x35S_annotated(fwd)          aaagatggacccccacccacgaggagcatcgtggaaaaagaagacgttccaaccacgtct   840
pGWB602omega_2x35S_annotated(fwd)    aaagatggacccccacccacgaggagcatcgtggaaaaagaagacgttccaaccacgtct   304
pCAMBIA2300_2x35S_annoted(fwd)       aaagatggacccccacccacgaggagcatcgtggaaaaagaagacgttccaaccacgtct   300
pCAMBIA1304_2x35S_annotated(fwd)     aaagatggacccccacccacgaggagcatcgtggaaaaagaagacgttccaaccacgtct   327
                                     ************************************************************

GB0222_2x35S_annotated(fwd)          tcaaagcaagtggattgatgtgatatctccactgacgtaagggatgacgcacaatc----   896
pGWB602omega_2x35S_annotated(fwd)    tcaaagcaagtggattgatgtgataacatggtggagcacgacactctcgtctactccaag   364
```

```
pCAMBIA2300_2x35S_annoted(fwd)          tcaaagcaagtggattgatgtgataacatggtggagcacgacactctcgtctactccaag   360
pCAMBIA1304_2x35S_annotated(fwd)        tcaaagcaagtggattgatgtgataacatggtggagcacgacactctcgtctactccaag   387
                                        ************************ *      **         *  **   * **

GB0222_2x35S_annotated(fwd)             ----------------------ccactatccttcgcaatgagacttttcaacaaagggta   934
pGWB602omega_2x35S_annotated(fwd)       aatatcaaagatacagtctcagaagaccaaagggctattgagacttttcaacaaagggta   424
pCAMBIA2300_2x35S_annoted(fwd)          aatatcaaagatacagtctcagaagaccaaagggctattgagacttttcaacaaagggta   420
pCAMBIA1304_2x35S_annotated(fwd)        aatatcaaagatacagtctcagaagaccaaagggctattgagacttttcaacaaagggta   447
                                                                          * **********************

GB0222_2x35S_annotated(fwd)             atatcgggaaacctcctcggattccattgcccagctatctgtcacttcatcaaaaggaca   994
pGWB602omega_2x35S_annotated(fwd)       atatcgggaaacctcctcggattccattgcccagctatctgtcacttcatcaaaaggaca   484
pCAMBIA2300_2x35S_annoted(fwd)          atatcgggaaacctcctcggattccattgcccagctatctgtcacttcatcaaaaggaca   480
pCAMBIA1304_2x35S_annotated(fwd)        atatcgggaaacctcctcggattccattgcccagctatctgtcacttcatcaaaaggaca   507
                                        ************************************************************

GB0222_2x35S_annotated(fwd)                   gtagaaaaggaaggtggcacctacaaatgccatcattgcgataaaggaaaggctatcgtt
         1054
pGWB602omega_2x35S_annotated(fwd)       gtagaaaaggaaggtggcacctacaaatgccatcattgcgataaaggaaaggctatcgtt   544
pCAMBIA2300_2x35S_annoted(fwd)          gtagaaaaggaaggtggcacctacaaatgccatcattgcgataaaggaaaggctatcgtt   540
pCAMBIA1304_2x35S_annotated(fwd)        gtagaaaaggaaggtggcacctacaaatgccatcattgcgataaaggaaaggctatcgtt   567
                                        ************************************************************

GB0222_2x35S_annotated(fwd)                   caagatgcctctgccgacagtggtcccaaagatggacccccacccacgaggagcatcgtg
         1114
pGWB602omega_2x35S_annotated(fwd)       caagatgcctctgccgacagtggtcccaaagatggacccccacccacgaggagcatcgtg   604
pCAMBIA2300_2x35S_annoted(fwd)          caagatgcctctgccgacagtggtcccaaagatggacccccacccacgaggagcatcgtg   600
pCAMBIA1304_2x35S_annotated(fwd)        caagatgcctctgccgacagtggtcccaaagatggacccccacccacgaggagcatcgtg   627
                                        ************************************************************

GB0222_2x35S_annotated(fwd)                   gaaaaagaagacgttccaaccacgtcttcaaagcaagtggattgatgtgatatctccact
         1174
pGWB602omega_2x35S_annotated(fwd)       gaaaaagaagacgttccaaccacgtcttcaaagcaagtggattgatgtgatatctccact   664
pCAMBIA2300_2x35S_annoted(fwd)          gaaaaagaagacgttccaaccacgtcttcaaagcaagtggattgatgtgatatctccact   660
pCAMBIA1304_2x35S_annotated(fwd)        gaaaaagaagacgttccaaccacgtcttcaaagcaagtggattgatgtgatatctccact   687
                                        ************************************************************

GB0222_2x35S_annotated(fwd)                   gacgtaagggatgacgcacaatcccactatccttcgcaagacccttcctctatataagga
         1234
pGWB602omega_2x35S_annotated(fwd)       gacgtaagggatgacgcacaatcccactatccttcgcaagacccttcctctatataagga   724
pCAMBIA2300_2x35S_annoted(fwd)          gacgtaagggatgacgcacaatcccactatccttcgcaagacc-ttcctctatataagga   719
pCAMBIA1304_2x35S_annotated(fwd)        gacgtaagggatgacgcacaatcccactatccttcgcaagacc-ttcctctatataagga   746
                                        ******************************************* ****************

GB0222_2x35S_annotated(fwd)                   agttcatttcatttggagaggactccggtatttttacaacaattaccacaacaaaacaaa
         1294
pGWB602omega_2x35S_annotated(fwd)       agttcatttcatttggagaggacacg----------------------------------   750
pCAMBIA2300_2x35S_annoted(fwd)          agttcatttcatttggagaggacacgctgaaatcaccagtctctctctacaaatctatct   779
pCAMBIA1304_2x35S_annotated(fwd)        agttcatttcatttggagaggacacgctgaaatcaccagtctctctctacaaatctatct   806
                                        ********************** *

GB0222_2x35S_annotated(fwd)             caacaaacaacattacaatttactattctagtcga         1329
pGWB602omega_2x35S_annotated(fwd)       ------------------------------------        750
pCAMBIA2300_2x35S_annoted(fwd)          ct----------------------------------        781
pCAMBIA1304_2x35S_annotated(fwd)        ct----------------------------------        808
```

A.IV BLAST to determine approximate transcript start and end for chlorovirus genes A208R and A067R

Figure A-1 Nucleotide BLAST of chlorovirus PBCV-1 A208R gene against RNA transcript data. The SRA archive was from the 20 minute timepoint from (Blanc et al., 2014) whose data was in experiment SRX317060. Green represents sequences with a lower alignment score. 1000 hits are shown.

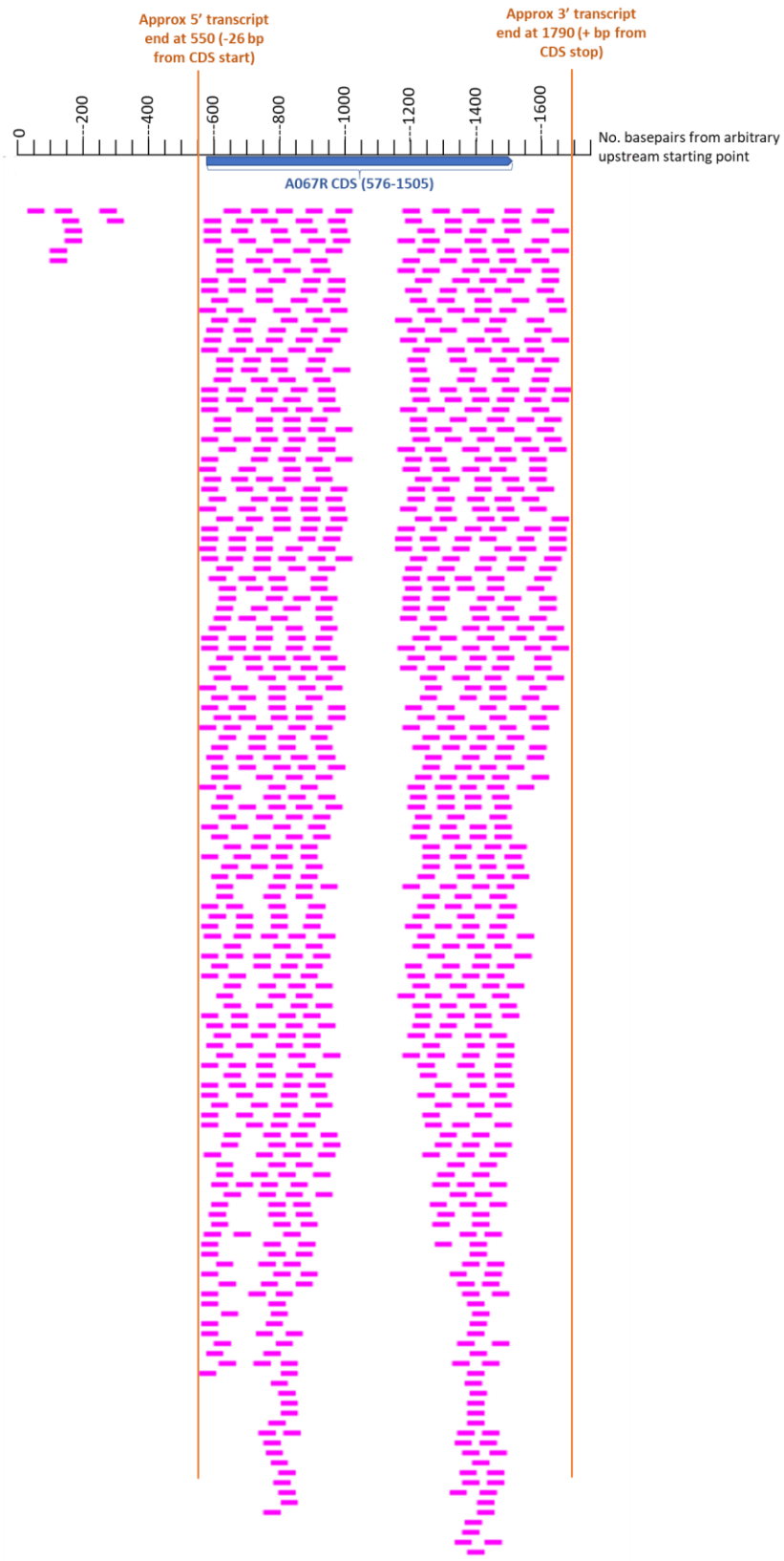


Figure A-2 Nucleotide BLAST of chlorovirus PBCV-1 A067R gene against RNA transcript data. The SRA archive was from the 20 minute timepoint from (Blanc et al., 2014) whose data was in experiment SRX317060. 1000 hits are shown.

A.V Plasmid Map

A.V.I pC-J23-div

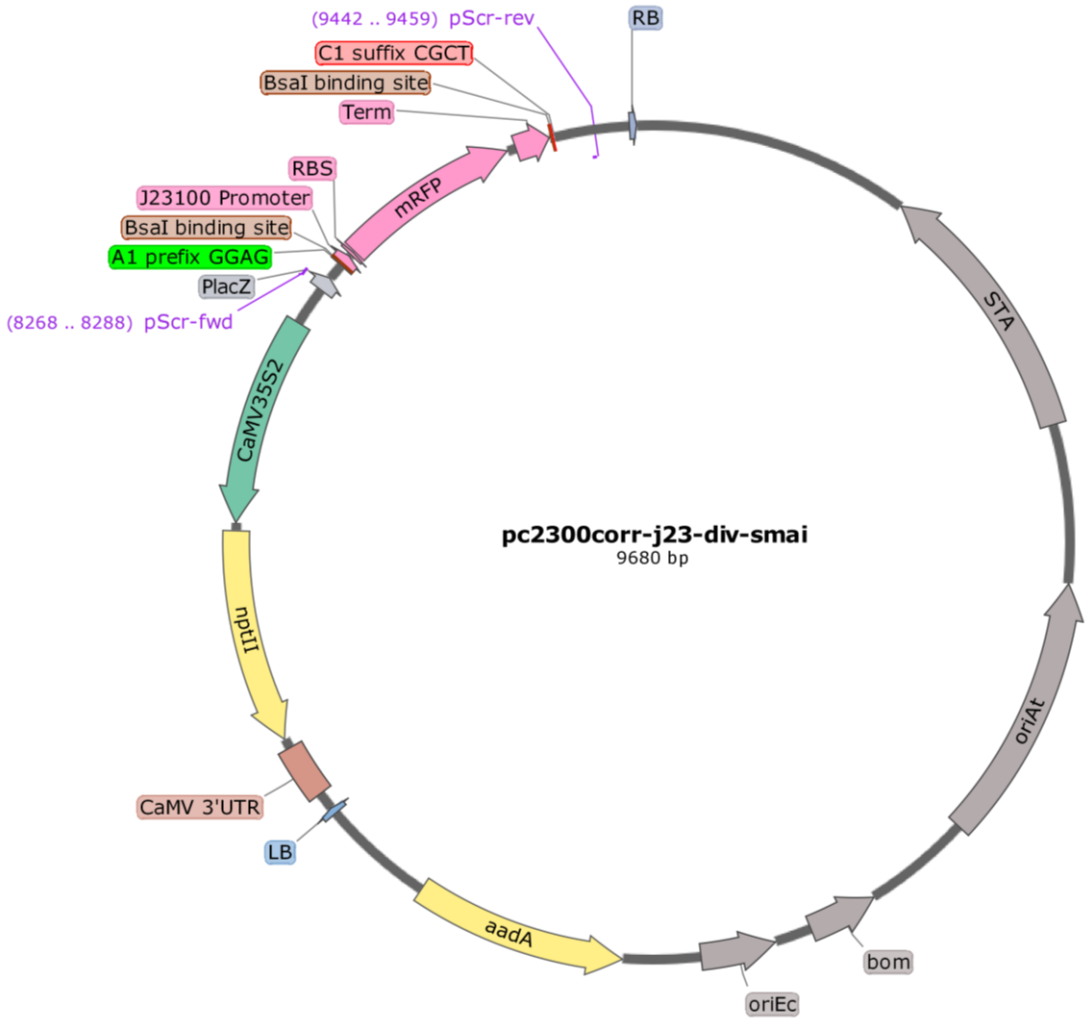


Figure A-3 Plasmid map of pC-J23-div, the Level 1 acceptor vector created and used in this work.

A.VI Golden Gate reaction plates and gels

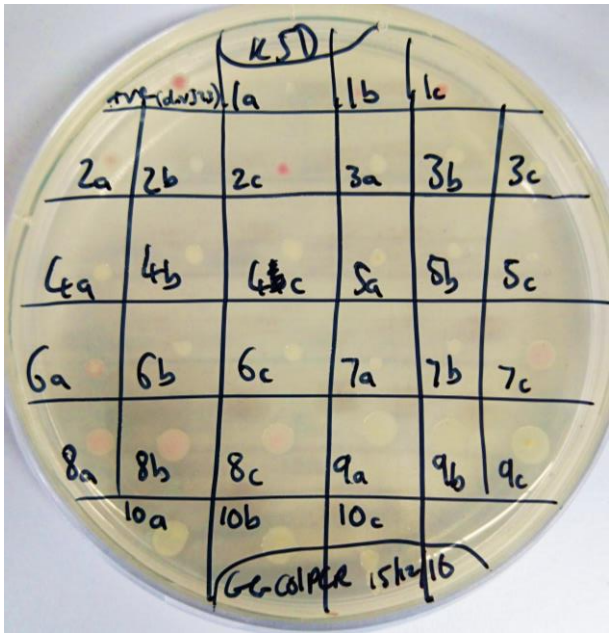


Figure A-4 Copy plate of picked colonies from GG1-10 assemblies. This data is shown in **Table 6-2**.

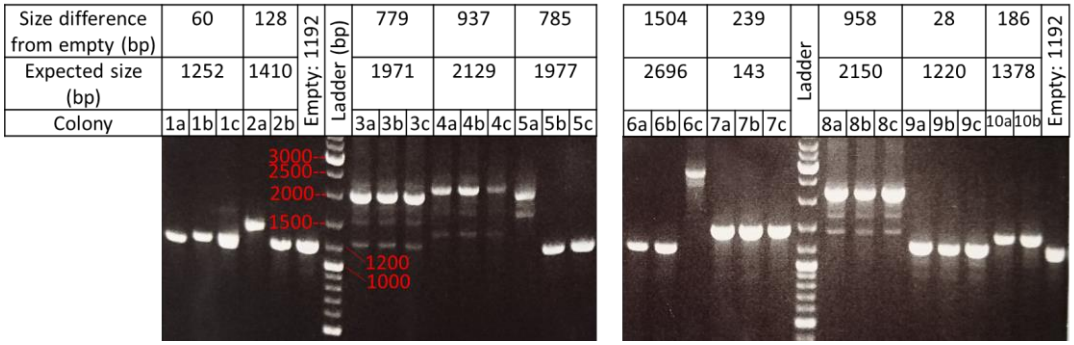


Figure A-5 Colony PCR screening of clones from GG1-10 assembly. This data is shown in **Table 6-2**.

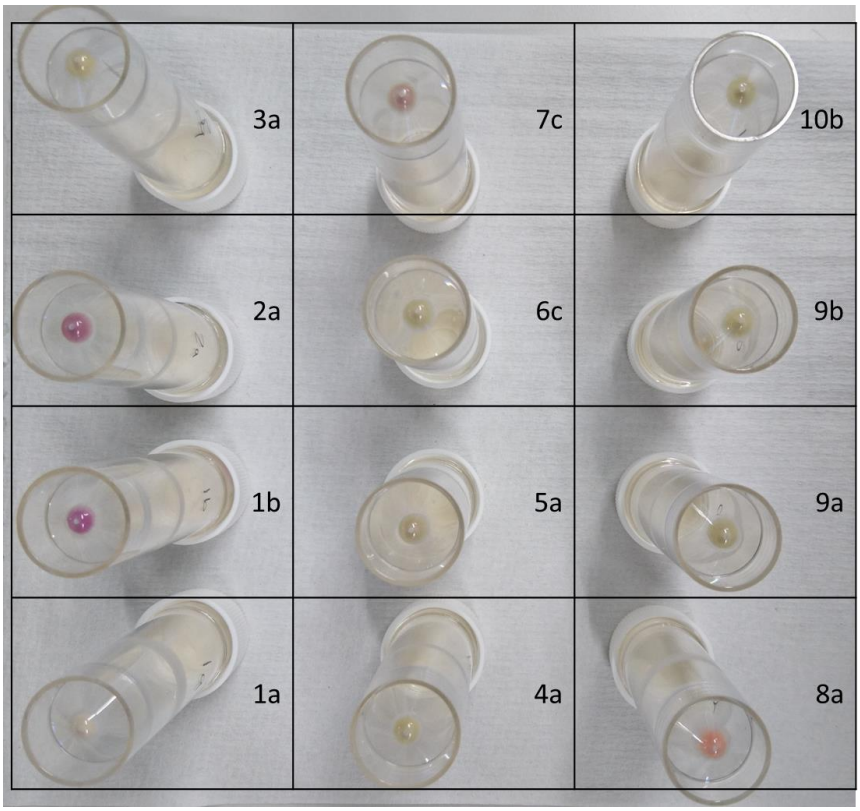


Figure A-6 Pellet colours from small overnight cultures of clones from GG1-10 assembly. This data is shown in **Table 6-2**.

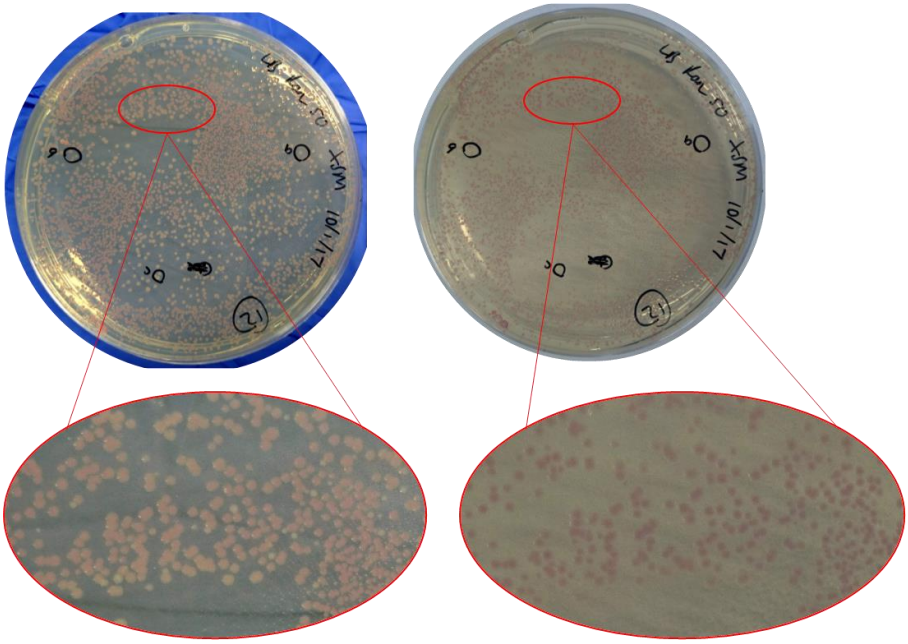
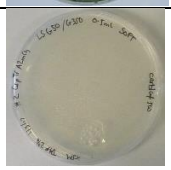
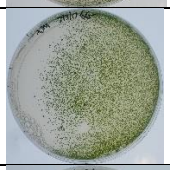
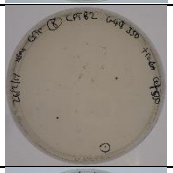
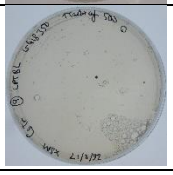


Figure A-7 Identification of positive and negative colonies from a Golden Gate assembly plate. Viewing plates against a blue background makes white positive colonies more easy to see than on a white/grey background. These images were taken after ~18-21 hours incubation. The assemblies were carried out using the old Golden Gate cycle (**Table 6-3**).

A.VII Selection plates and propagations from *Chlorella* transformations with Golden Gate constructs via *Agrobacterium*-mediated transformation

ID	Description	DSS 15th day growth	LSS 10th day growth
Rb	RbleC23 +ve control		
GG1	A2-mCh		 Note: 12th day growth for this plate (no colonies)
GG2	A6-mCh		
GG5	CPT-mCh		
GG7	RPT-mCh		
GG12	RPT-B2		
GG16	RPT-BL		
GG11	CPT-B2		
GG15	CPT-BL		

-veAg	Co-incubation was performed with untransformed *Agrobacterium*		
-ve	Nothing was co-incubated with *Chlorella*		

Table A-6 Primary selection plates from an *Agrobacterium*-mediated transformation of *Chlorella sorokiniana*. Selection was performed at 300-350 µg/ml G418, with (LSS) or without (DSS) a 2-5 day liquid enrichment step of 50 µg/ml G418.

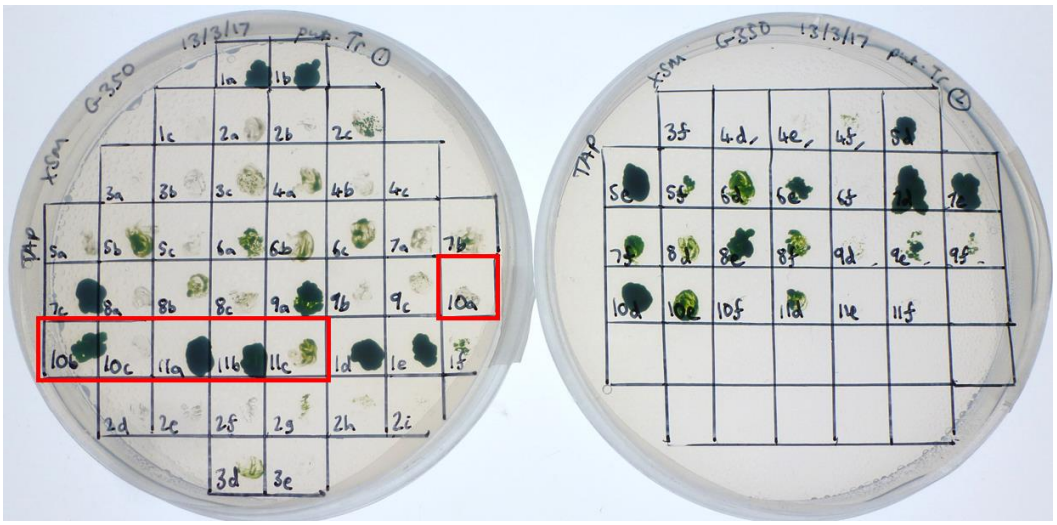


Figure A-8 Propagation plate of putative *Chlorella sorokiniana* transformants. The antibiotic concentration is 350 µg/ml. The primary selection plates are shown in **Table A-6**. The red boxes highlight colonies isolated from negative control plates. The growth of these occurring at equivalent or better growth than the other picket colonies suggests that the others are false positives and hence analysis was discontinued for these clones.